EPIGENETIC REGULATION OF DNA REPAIR MEDIATED BY THE HISTONE METHYLTRANSFERASE DOT1L

BY

Nehemiah Seth Alvarez

Submitted to the graduate degree program in Pathology and Laboratory Medicine and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

_____

Chairperson: Dr. Patrick E. Fields

_____

Dr. Fariba Behbod

_____

Dr. Timothy A. Fields

_____

Dr. Kenneth R. Peterson

_____

Dr. Joseph D. Fontes

Date Defended: November 28th, 2016

The Dissertation Committee for Nehemiah Seth Alvarez certifies that this is the approved version of the following dissertation:

EPIGENETIC REGULATION OF DNA REPAIR MEDIATED BY THE HISTONE METHYLTRANSFERASE DOT1L

_____

Chairperson: Dr. Patrick E. Fields

Date approved: November 29th, 2016

Abstract

The accurate transmission of a genome into future generations, otherwise known as genome integrity, is necessary for life. However, genomic changes due to error or damage are quite common, with an estimated 38,000 DNA base alterations within a cell at any given moment. Thus, genomic integrity depends on high fidelity repair of DNA damage caused by endogenous and exogenous sources. Every cell has machinery that is capable of repairing DNA lesions, and defects in genes encoding DNA repair machinery frequently result in debilitating developmental disorders, cancer, and/or death.

There are seven main pathways for DNA repair that have been identified: base excision repair, mismatch repair, nucleotide excision repair, interstrand crosslink repair, single strand break repair, non-homologous end joining, and homologous recombination. Pathway utilization depends on both the type of damage and the stage of the cell cycle. Homologous recombination is an especially important pathway, as it is the sole high fidelity means of repairing double stranded DNA breaks, which are particularly deleterious. It is estimated that ~10 unrepaired double strand breaks are lethal to a cell. The mechanisms that control double strand break repair are incompletely understood, but histone variant H2AZ exchange at DNA damage sites is believed to be an early step in the process and may be regulated by core histone modifications.

DOT1-like (*DOT1L*), the human ortholog of the yeast gene DOT1 (Disruptor of telomeric silencing-1), is a histone H3 (H3K79) methyltransferase and has been implicated in DNA repair. This volume describes the use of both biological and computer science-based approaches to examine the specific function of DOT1L in DNA repair. Cells from DOT1L knockout mice were found to exhibit marked genomic instability and defects in DNA repair. Further, loss of DOT1L and/or its methylation activity resulted in decreased H2AZ incorporation

at double strand breaks and increased amounts of single strand DNA, leading to a specific defect in homologous recombination repair activity. Instead, in the absence of either DOT1L or H3K79 methyltransferase activity, the non-homologous end joining repair pathway was utilized, resulting in genomic instability. These findings identify a novel role for DOT1L and H3K79 methylation in facilitating histone exchange at double strand break sites, generating a chromatin state that is permissive for homologous recombination repair. This discovery emphasizes the diverse roles of epigenetic modification in biological processes beyond transcriptional control. Moreover, the identification of DOT1L function in DNA repair provides new opportunities for design and implementation of molecular therapeutics strategies that target DNA repair pathways to treat diseases.

Dedication

To all those that come after; I hope you can see further into the unknown than I have.

Acknowledgments

I would like to thank my Ph.D. advisor, Dr. Patrick E. Fields, for allowing me the opportunity to work in his laboratory and work on DNA damage repair. Dr. Fields encouraged me to pursue novel avenues of research, yet kept my work on a directed path. I would also like to thank my committee members, Dr. Fariba Behbod, Dr. Timothy A. Fields, Dr. Kenneth R. Peterson, and Dr. Joseph D. Fontes. They provided me with direction and experimental insight. Most importantly, I would like to thank my family who has given me unwavering support. Lastly, I would like to sincerely thank Dr. Pavla Brachova. This research would not have been possible without her continuous support and encouragement.

Table of Contents

List of Tables and Figures

List of Abbreviations

53BP1 - tumor protein p53 binding protein 1
altNHEJ - alternative nonhomologous end joining
APE1 - DNA-(apurinic or apyrimidinic site) lyase
APLF - aprataxin-PNK–like factor
ATM - Ataxia telangiectasia mutated
BEL-ChIP - broken end ligation chromatin immunoprecipitation
BER - base excision repair
BLM - Bloom syndrome RecQ like helicase
BRCA1 - breast cancer 1
BRCA2 - breast cancer 2
BRG1 - switch/sucrose non-fermentable (SWI/SNF) Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4
BRIP1 - breast cancer 1 (BRCA1) interacting protein C-terminal helicase 1
cNHEJ - classical nonhomologous end joining
CSA - Cockayne syndrome WD repeat protein
CSB - Cockayne syndrome protein B
CtIP - MRN-BRCA1/retinoblastoma binding protein 8
DDB1 - damage specific DNA binding protein 1
DDB2 - DNA binding protein 2
DNA - deoxyribonucleic acid
DNA-PKcs - DNA-activated, catalytic polypeptide
DOT1L - disruptor of telomeric silencing 1-Like
DSS1 - split hand/foot malformation type 1
DSB – double strand break
ERCC6 - Cockayne syndrome protein B
ERCC8 - Cockayne syndrome WD repeat protein
FAN1 - FANCD2/FANCI-associated nuclease 1
FANCA, B, C, G, E, L – Fanconi anemia associated genes
G1 - Gap 1 phase of cell cycle
G2 - Gap 1 phase of cell cycle
ggNER - global genome nucleotide excision repair
γH2AX – gamma H2A histone family member X
H1, 2, 3, 4 – histone 1, 2, 3, 4
H2AZ - histone 2A family member Z
H3K - histone 3 lysine
HPF1 - histone PARylation factor 1
HR - homologous recombination
Htz1 - yeast homologue of mammalian H2AZ, histone 2A Z1
ICLR - interstrand crosslink repair
INO80 - INO80 complex subunit
ISW1 - Imitation-switch 1
ISW2 - Imitation-switch 2
JMJD1C - jumonji domain containing 1C
JMJD2A - jumonji domain containing 2A

KU70/80 – XRCC6/XRCC5
LIG1 - DNA ligase I
LIG3 - DNA ligase III
M - mitotic phase of cell cycle
MANA - Machine autonomous nuclei analyzer
MDC1 - mediator of DNA damage checkpoint 1
MHF1 - centromere protein S
MHF2 - centromere protein X
MMEJ - microhomology mediated end joining
MMR - mismatach repair
MRE11 - Meiotic Recombination 11 homolog
MRN - complex composed of Meiotic Recombination 11 homolog (MRE11)/radiation sensitive 50 (RAD50)/NLR family pyrin domain containing 2 (NLRP2/NBS1)
MSH2 - methyl-directed mismatch repair (mutS) homolog 2
MSH3 - mutS homolog 3
MSH6 - mutS homolog 6
MSL1 - male-specific lethal 1 homolog
NBS1 - NLR family pyrin domain containing 2
NER - nucleotide excision repair
NHEJ - nonhomologous end joining
OGG1 - 8-Oxoguanine glycosylase
PALB2 - partner and localizer of BRCA2
PARP - Poly(ADP-ribose) polymerase 1
PCNA - proliferating cell nuclear antigen
PMS2 - postmeiotic segregation increased-1 homolog 2
PNKP - polynucleotide kinase 3'-phosphatase
RAD23B, 50, 51C, 52/51 - radiation sensitive genes
RAP80 - ubiquitin interaction motif containing 1
RB1 - retinoblastoma transcriptional corepressor 1
RMI - RecQ-mediated genome instability complex
RNA - ribonucleic acid
RNF8 - ring finger protein 8
ROS - reactive oxygen species
RPA - replication protein A
S – synthesis phase of cell cycle
SLX1 - synthetic Lethal of unknown (X) function homolog 1
SLX4 - synthetic Lethal of unknown (X) function homolog 4
SNF - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4provided
SSB1 - single stranded binding homologue 1
SSBR - single strand break repair
SSB – single strand break
ssDNA - single stranded DNA
SWI - switch/sucrose non-fermentable
tcNER - transcription coupled nucleotide excision repair
TDP1 - Tyrosyl-DNA phosphodiesterase-1

Tip60 - lysine acetyltransferase 5
TOP1 - Topoisomerase I
TOPIIIa - Topoisomerase (DNA) III Alpha
UBC13 - E2 ubiquitin-conjugating protein 13
UIMC1 - ubiquitin interaction motif containing 1
USP7 - ubiquitin-specific-processing protease 7
UV – ultraviolet
UVSSA - UV-stimulated scaffold protein A
XAB2 - XPA-binding protein 2
XPC, F, G - xeroderma pigmentosum genes
XRCC1 - x-ray repair cross-complementing protein 1
XRCC6 - X-ray repair cross-complementing protein 6

Chapter I: Introduction

How Is Genome Stability Maintained?

Genome stability is a cellular trait that ensures accurate propagation of DNA. In order for a cell to maintain genome stability, it must be able to repair DNA damage induced by environmental changes. DNA damage can be induced by a variety of phenomenon such as deamination, depurination, alkylation, replication induced nucleotide misincorporation, and reactive oxygen species that are generated from cellular metabolism (Table 1.1). (ref for the table?)

Lesion Type	Cause	Frequency/Per Cell
DNA base alterations	CM	38,118
DNA base misincorporation	DR/MT	$< 10^{-5}$
Pyrimidine/Thymine dimers	UV	10,000
Intrastrand crosslinks	IR/MT	?
Single strand breaks	IR/MT	1,000
Double strand breaks	IR/MT	2

Table 1.1: Summary of DNA lesion type, cause, and frequency. (CM) cellular metabolism, (DR) DNA replication, (MT) molecular therapy, (IR) ionizing radiation. The frequency represents the average DNA lesion per cell at any given time.

It is estimated that a cell can have more than an average of 38,000 DNA base alterations at any one time caused by endogenous DNA damaging factors [1]. Nucleotide misincorporation due to DNA replication is estimated to be less than 10^{-5} in normal tissue, but the vast majority of these alterations are repaired by cellular DNA repair machinery [2]. However, in disease states such as cancer, nucleotide misincorportation can be high due to mutations in genes required for DNA repair [3]. Other drivers of DNA damage include ionizing and ultraviolet radiation. Exposure to environmental ionizing radiation from sources such as radon (Rn-222) varies by

geographical location, but can generate two double strand breaks and over 1000 single strand breaks per cell [4]. UV radiation from direct sunlight can generate up to 10,000 lesions per cell, depending on the level of exposure [4].

Molecular therapies used to treat cancer can also cause DNA damage. Four broad groups of chemotherapies exist that induce DNA damage: alkylating agents, antimetabolites, topoisomerase inhibitors, and cytotoxic antibiotics. Alkylating agents such as mechlorethamine, N-Nitroso-N-methylurea, dacarbazine, and cisplatin introduce covalent bonds between DNA and the proteins that bind them. Antimetabolites inhibit DNA or RNA synthesis. Treatment with antimetabolites such as anti-folates or thiopurines can induce DNA damage through the misincorporation of bases. The use of topoisomerase inhibitors such as irinotecan etoposide, and novobiocin act to prevent the release of tension produced on DNA through replication and transcription. The tension on DNA induced by replication and transcription is enough to break DNA if it is not released by the action of topoisomerase. Cytotoxic antibiotics such as doxorubicin, and bleomycin introduce DNA damage through DNA intercalation and free radical generation that can cause double strand breaks.

Cells have evolved mechanisms to respond to environmental challenges that damage DNA through the evolution of specialized DNA repair pathways. Chemical alterations on DNA bases are replaced via base excision repair (BER) pathway, while the misincorporation of bases is corrected by mismatch repair (MMR) pathway. Environmental challenges that introduce pyrimidine and thymine dimers are repaired by nucleotide excision repair (NER) pathway. Interstrand crosslinks are repaired via the interstrand crosslink repair (ICLR) pathway, which shares components with other DNA repair pathways. The single strand break repair (SSBR)

pathway repairs single strand breaks, while double strand breaks are repaired via homologous recombination (HR) or non-homologous end joining (NHEJ) repair pathways.

The use of individual repair pathways not only depends on the type of damage, but also on the stage of the cell cycle in which the damage was encountered (Table 1.2) [5-10]..

Lesion Type	G1	S	G2/M
DNA base alterations	BER		
DNA base misincorporation		MMR	
Pyrimidine/Thymine dimers	NER	NER	NER
Intrastrand crosslinks		ICLR	
Single strand breaks	SSBR	SSBR	SSBR
Double strand breaks	NHEJ/HR	NHEJ	NHEJ/HR

Table 1.2: Summary of DNA lesion type and the phase of the cell cycle of maximum activity. (BER) base excision repair, (MMR) mismatch repair, (NER) nucleotide excision repair, (ICLR) interstrand crosslink repair, (SSBR) single strand break repair, (NHEJ) nonhomologous end joining, (HR) homologous recombination. G1/S/G2/M are the phases of the cell cycle.

Components of BER pathway are most active in the G1 phase of the cell cycle, however the repair can take place in other cell cycle phases as well [5]. The MMR pathway has maximal activity during S phase [6]. NER repair is active throughout the cell cycle [7]. ICLR pathway is most active during S phase [8]. SSB repair is hypothesized to be active at all phases of the cell cycle [9]. HR repair is maximally active in the S phase , while and NHEJ repair is active during all phases of the cell cycle [10].

DNA repair is carried out by multitude of proteins that modify DNA and the proteins surrounding the break site. The classes of proteins involved include kinases, acetyltransferases, methyltransferases, glycosylases, ubiquitin ligases, DNA ligases, helicases, polymerases, topoisomerases, recombinase, demethylases, deacetylases, phosphatases, and chaperons. The

spatial and temporal assembly of DNA repair factors must be tightly controlled in order to maintain genome stability.

The complexity of the various DNA repair pathways is built upon two simple processes: detecting the DNA damage lesion and recruiting factors to the site of DNA damage. In eukaryotic cells genomic DNA is bound by histones and these histones serve as a binding platform for various DNA repair factors. Histones bind together with DNA to form a nucleosome. The nucleosome is composed of an H2A-H2B tetramer and an H3-H4 tetramer and DNA is wrapped around the entire complex. The nucleosomes can protect DNA from sources of endogenous DNA damage by limiting the surface area of DNA exposed to genotoxic agents. The intervening sequence between nucleosomes is more vulnerable to for DNA damaging agents.

Histones have four major types of modifications that constitute the histone code: acetylation, phosphorylation, methylation, and ubiquitination. Components of DNA repair pathways read the histone code in order to initiate specific types of DNA repair. The continual presence of the histone code serves as a marker that instructs the DNA repair machinery on the type of repair the needs to take place at specific regions of the genome.

The following sections will outline details of the various DNA repair pathways. The response to DNA damage is as diverse as the types of genotoxic factors cells are exposed to. Furthermore, this review will present the idea that utilization of specific repair pathways is governed by the epigenetic state of chromatin that is present in the region damaged. We hypothesize that the epigenetic landscape primes regions of the genome for different types of DNA repair. This hypothesis has broad implications, in particular for the administration of molecular therapies for cancers that target epigenetic factors.

Base excision repair

The base excision repair pathway can repair chemical alterations on DNA bases. The initial sensing of DNA base alterations is performed by DNA glycosylases. There are 11 unique DNA glycosylases in humans, each responsible for detecting a different type of DNA base modification [11]. The mechanism by which these glycosylases can detect a modified DNA base in the genome is still unclear. However, the human 8-Oxoguanine glycosylase (OGG1) can actively scan the DNA in a bilateral fashion looking for abnormal bases [12](Figure 1.1A). When a glycosylase encounters a modified base, it is excised leaving an abasic site (Figure 1.1B). The abasic site is a target for the DNA-(apurinic or apyrimidinic site) lyase (APE1), which cleaves 5′ of the abasic site generating a gap in the DNA [13] (Figure 1.1C). The gap in the DNA can be short (a single base) or long (2-10 bases). Depending on the length of the gap and the type of glycosylase that initiated the repair, different proteins are used to fill in the missing nucleotides [14]. Short-patch base excision utilizes DNA polymerase β (Pol β), DNA ligase I/ III (LIG1/LIG3), and x-ray repair cross-complementing protein 1 (XRCC1) to fill the gap [15](Figure 1.1D). Long-patch base excision repair takes place in proliferating cells and utilizes components of the replication machinery: Proliferating cell nuclear antigen (PCNA), DNA Ligase 1 (LIG1), and DNA polymerase ∂/ε (Pol ∂/Pol ε)[16] (Figure 1.1E).

There is evidence indicating that chromatin remodeling is required for efficient base excision repair processes. Studies on OGG1 show that removal of histone 1 (H1) is required for repair of modified bases [17]. Base excision repair using DNA polymerase β requires an active Imitation-switch 1 (ISW1) and Imitation-switch 2 (ISW2) remolding complexes [18]. These studies suggest that the chromatin state can influence repair fidelity at site of DNA damage. It stands to reason that factors, which alter chromatin remolding, could modulate BER activity. An interesting approach would be to test molecular therapeutics that target epigenetic marks in

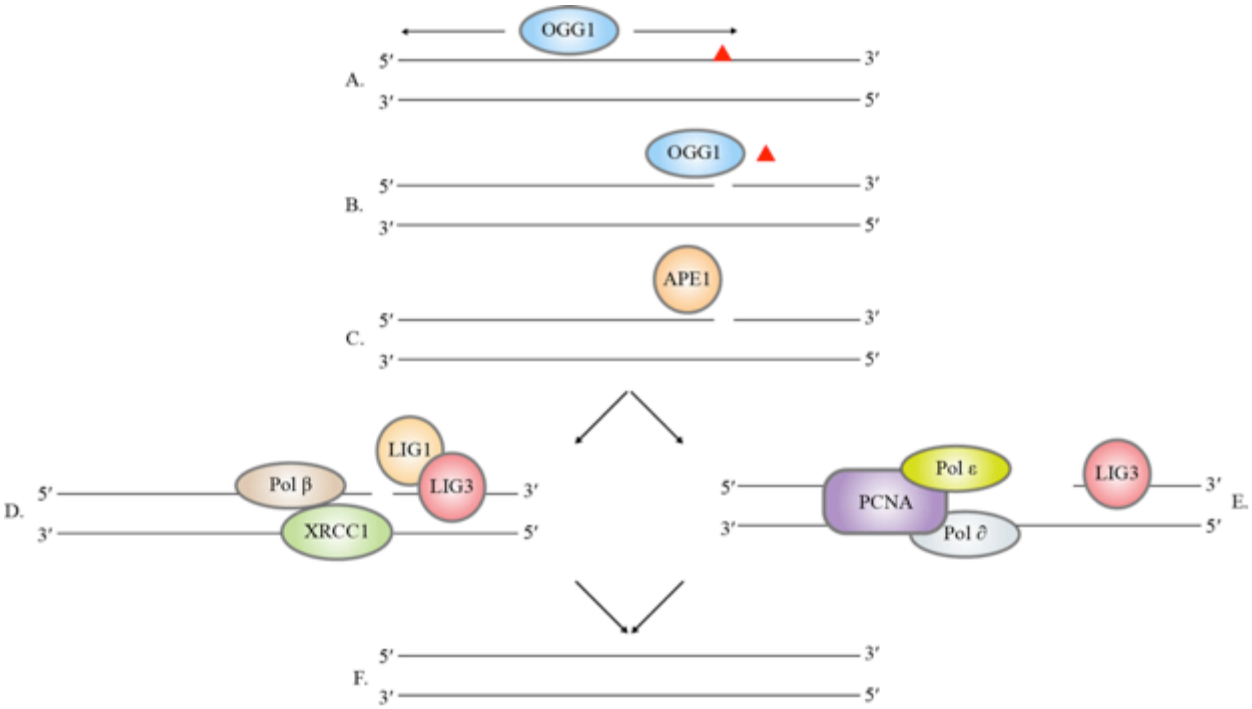


Figure 1.1: Base excision repair pathway. A. OGG1 scans DNA looking for modified bases. B. Upon recognition of a modified base OGG1 excises the base. C. APE1 is recruited to the abasic site and cleaves five prime of the abasic site. D. If the cleavage 1 base, then short patch repair is used. E. If the cleaves is over 2 bases long patch repair is used. F. DNA ligase seal the filled in strands to generate double stranded product.

conjunction with methanesulfonate, a highly mutagenic compound, to see if, similar to Poly(ADP-ribose) polymerase 1 (PARP) inhibitors, they can enhance or restrict base excision repair.

Mismatch repair

The mismatch repair pathway is activated by the misincorporation of bases. Base mismatch errors are the least encountered type of DNA lesion during normal physiological conditions. This is due to the high fidelity proof reading capability of DNA polymerases that keep the numbers of mismatched bases low, and the activity of the MMR machinery itself. Incorporation of an incorrect base by the replication machinery is estimated to be less than 10^{-5}

per base, and the MMR machinery has a high correction rate for most types of mismatches encountered that escape the proofreading capacity of DNA polymerases [19].
The MMR machinery also processes insertion and deletion loops (IDLs), which are a result of primer strand slippage during replication [20].

The principal components of the MMR machinery are male-specific lethal 1 homolog (MSL1), methyl-directed mismatch repair (mutS) homolog 2 (MSH2), mutS homolog 3 (MSH3), mutS homolog 6 (MSH6), postmeiotic segregation increased-1 homolog 2 (PMS2), proliferating cell nuclear antigen (PCNA), replication factor C (RFC), DNA polymerase δ (Pol δ) and DNA ligase I (LIG1), replication protein A (RPA), and exonuclease-1 (EXO1). These factors associate together in various complexes that repair specific types of DNA mismatches. The heterodimers MSH2/MSH6 form a complex called MutSα and repair single base mismatches and 1-2 IDL base mismatches, MLH1/PMS2 form the MutLα complex that associates with MutSα during repair, while MSH2/MSH3 form a complex called MutSβ and primarily repair IDL of greater than 2 bases, but how it functions is still currently unknown [21].

In humans, MutSα and MutLα recognize base mismatches by detecting the distortion in the backbone of DNA caused by the mismatched bases while scanning the DNA [22](Figure 1.2A). Upon binding MutSα/MutLα undergo translocation in either direction until they encounter and displace RFC from PCNA [23](Figure 1.2B). The displacement of RFC triggers the recruitment of EXO1, which degrades in the 5′→3′ direction [24](Figure 1.2C). The ssDNA that is generated is coated by RPA [25](Figure 1.2C). It is hypothesized that EXO1, PCNA and

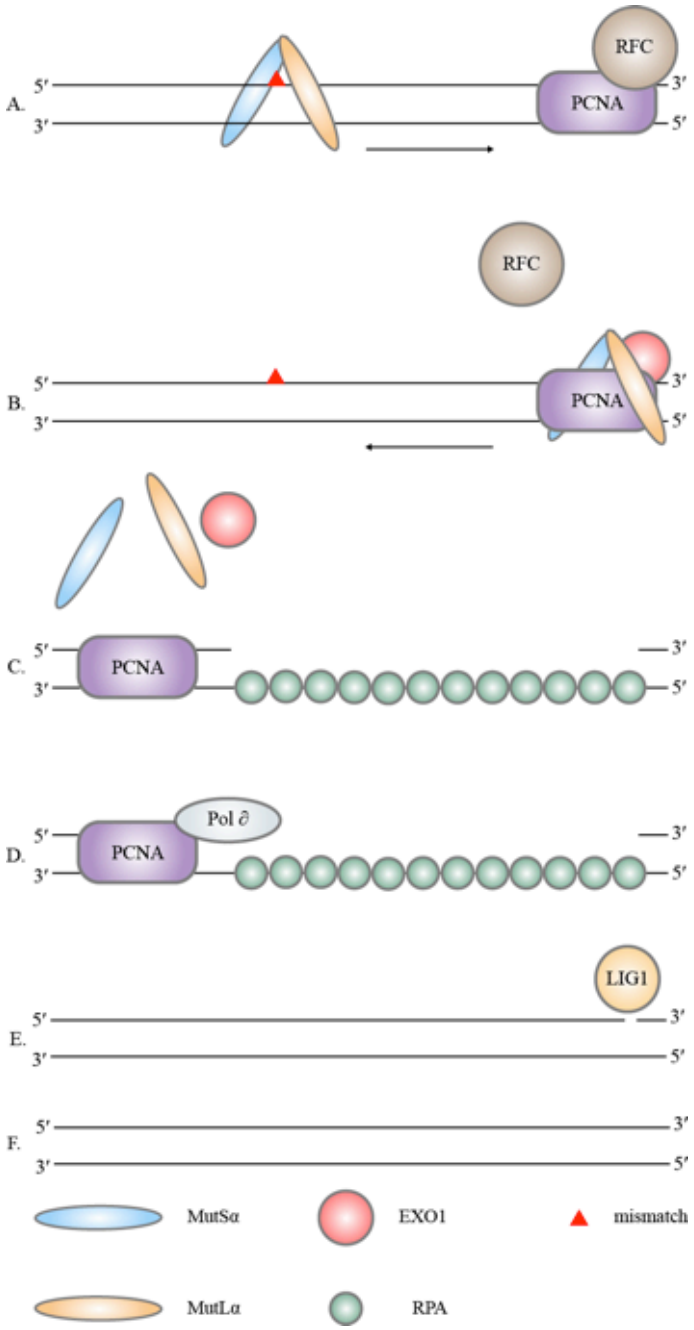


Figure 1.2: Mismatch repair pathway. A. MutSα/MutLα detect the mismatch and translocate to PCNA. B. MutSα/MutLα/PCNA recruit EXO1 which degrades the DNA in the 5′→3′ direction, while RPA coats the ssDNA. C. The degradation of the mismatched base cases MutSα/MutLα and EXO1 to dissociate from PNCA. D. PCNA recruits Pol δ to fill in the gap and (E.) LIG1 seals the strand producing (F.) a double stranded product.

MutSα/MutLα move together as a complex and once the mismatch is removed this leads to destabilization of the EXO1/MutSα/MutLα complex (Figure 1.2C) and recruitment of Pol δ and LIG1 to seal the gap [26] (Figure 1.2D/F/G).

The traversing of MMR complex raises the question of how MMR functions in the presence of nucleosomes. It was revealed that the binding of MSH2/MSH6 to DNA results in chromatin remodeling around the mismatched base [27]. The binding of MSH2/MSH6 results in nucleosome disassembly, and this effect is enhanced when a modified H3K4 acetylation mimic was used [27]. This is intriguing because it suggests that histone modifications can modulate MMR. It has been demonstrated that cells lacking H3K36 tri-methylation have an identical mutational pattern to cells lacking a functional MMR pathway [28]. With this in mind, it would be interesting to perform a genome-wide assessment to test for a correlation between H3K36 methylation and MMR hot spots.

Nucleotide excision repair

Nucleotide excision repair can fix a wide variety of DNA lesions such as pyrimidine dimers and thymine dimers, cyclopurines, and intrastrand crosslinks that are introduced by UV radiation. There are two sub-pathways of NER: global genome nucleotide excision repair (ggNER) and transcriptional coupled nucleotide excision repair (tcNER). ggNER scans the entire genome looking for strand abnormalities, while tcNER is activated by template strand lesions that stall RNA polymerase II.

The main damage sensor for ggNER is the xeroderma pigmentosum C (XPC)/radiation sensitive 23 B (RAD23B)/centrin 2 (CETN2) complex, which actively searches DNA for mutations [29]. DNA damage caused by UV radiation decreases the affinity of

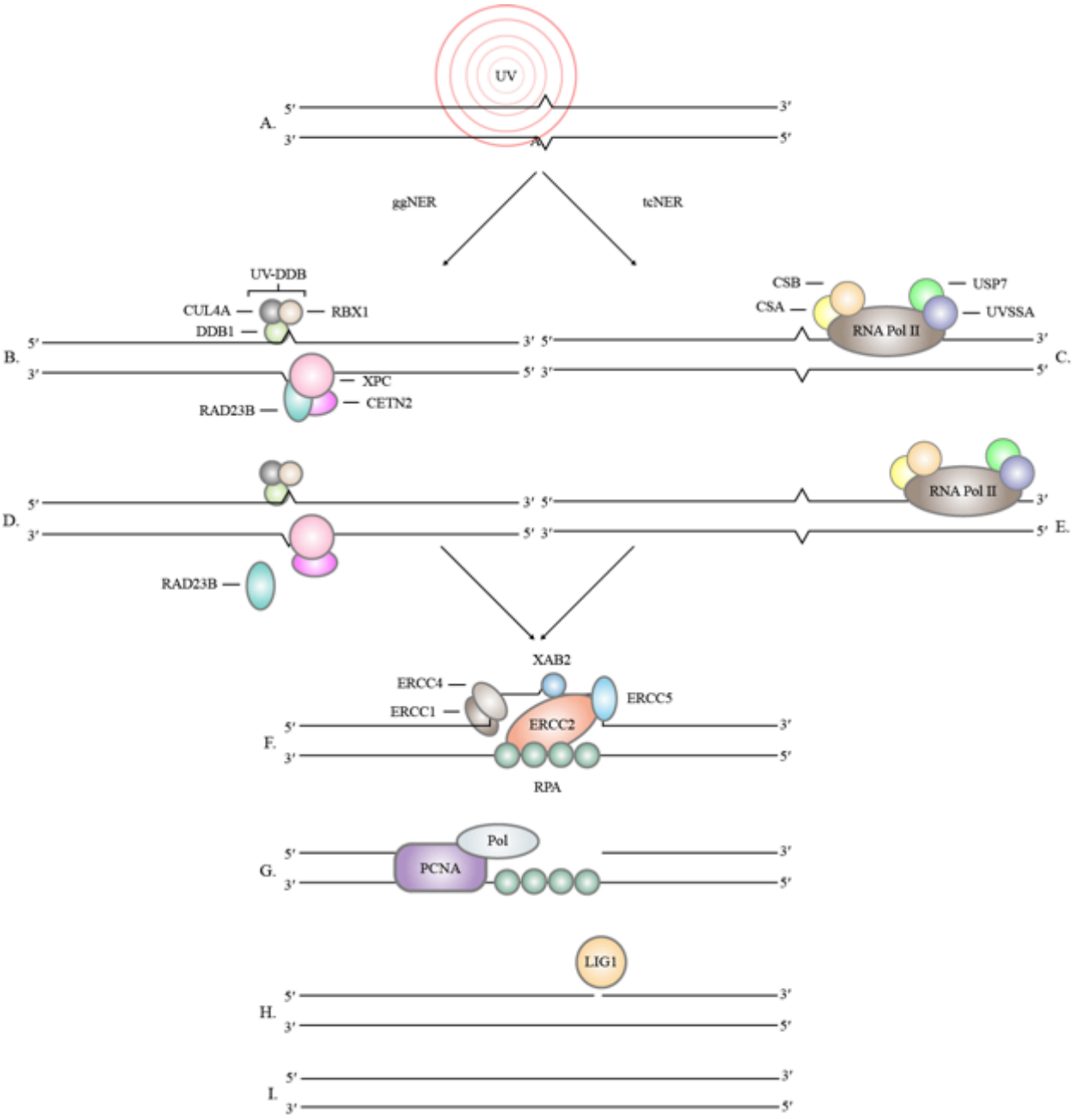


Figure 1.3: Nucleotide excision repair pathway. A. UV radiation introduces a DNA lesion. B. During ggNER, XPC/RAD23B/CETN2 and UV-DDB detect and bind to the lesion. C. In tcNER, CSA/CSB recruit UPS7/UVSSA to stalled RNA polymerase II and cause it to (E.) back track away from the lesion. D. During ggNER UV-DDB causes the dissociation of RAD23B from XPC/CETN2. F. ERCC2 is recruited to the lesion and unwinds the DNA allowing ERCC4/ERCC1 access to the 5′ end of the lesion and ERCC5 access to the 3′ end. ERCC4/ERCC1 and ERCC5 cut the DNA excising the damaged portion. G. PCNA and one of three DNA polymerases (Pol δ, Pol ε, Pol κ) are recruited to the site to fill in the gap. H. DNA Ligase I seals the strand (I.) generating a dsDNA product.

XPC/RAD23B/CENT2 complex, allowing the UV-DDB complex composed of ubiquitin factors damage specific DNA binding protein 1 (DDB1), cullin 4A (CUL4A), ring-box 1 (RBX1), and DNA binding protein 1 (DDB2) to bind to the DNA [30](Figure 1.3B). The binding of UV-DDB causes RAD23B to dissociate from the UV-DDB complex, and initiates the recruitment of the NER repair machinery (Figure 1.3D).

During tcNER, the damage is detected indirectly through RNA polymerase II stalling. Stalling of RNA polymerase leads to the recruitment of the Cockayne syndrome WD repeat protein (CSA, ERCC8) and Cockayne syndrome protein B (CSB, ERCC6) [31]. When CSA and CSB bind the RNA polymerase II, this results in the recruitment of a complex of proteins (UV-stimulated scaffold protein A (UVSSA), ubiquitin-specific-processing protease 7 (USP7) (Figure 1.3C) and causes RNA polymerase II to back track [29, 32](Figure 1.3E).

At this point, the tcNER and ggNER pathways converge and both tcNER and ggNER complex can recruit XPA-binding protein 2 (XAB2), excision repair cross-complementing 2 (ERCC2) and the excision repair cross-complementing 5 (ERCC5) endonuclease [29](Figure 1.3F). The helicase activity of ERCC2 unwinds the DNA around the lesion, allowing the excision repair cross-complementing 4 (ERCC4)/excision repair cross-complementing 1 (ERCC1) complex to cut the DNA 5′ of the lesion and ERCC5 to a cut 3′ of the lesion, and thus excise the damaged section of DNA [29](Figure 1.3F). Finally, PCNA, and one of three polymerases, Pol δ, DNA polymerase ε (Pol ε), DNA polymerase (Pol κ) and LIG1 are recruited to fill in and seal the gap (Figure 1.3G,H).

Chromatin remolding is required for both ggNER and tcNER to take place. The Switch/Sucrose non-fermentable (SWI/SNF) chromosome remodeling complex and INO80 complex facilitate NER by recruiting XPC and XPA to sites of damage [33, 34]. Interestingly it

has been reported that disruptor of telomeric silencing 1-Like (DOT1L) functions in NER [35]. However, the authors attribute the effects of DOT1L on NER to transcriptional regulation rather than direct effects. Results presented in Chapter V encourage a revisit to the understanding of the role of DOT1L in NER. We show that DOT1L mutants have profound defects in histone remodeling around sites of DNA damage (Chapter II/IV). Specifically, we observe strong defects in histone 2A family member Z (H2AZ) incorporation following DNA damage induction in DOT1L deficient cells. It is of note that the yeast homologue of mammalian H2AZ, histone 2A Z1 (Htz1), is a positive regulator of NER and its incorporation into nucleosomes facilitates efficient NER. It would be interesting to test if the role of DOT1L in NER is a result of defects in H2AZ incorporation at sites of DNA damage, as has been observed for double strand break repair (Chapter II/V).

Interstrand Crosslink Repair

Interstrand crosslinks of DNA are particularly toxic to cells because the separation of DNA is prevented, thereby inhibiting a number of necessary cellular processes such as DNA replication, transcription, and DNA repair. Thus, interstrand crosslinking agents such as cisplatin and nitrogen mustards are a common strategy for anti-cancer therapy. The formation of interstrand crosslinks is not solely due to human intervention, however. Exposure to environmental stresses can lead to their formation, and organisms evolved strategies to counteract their formation. The Fanconi anemia pathway is the primary pathway used for repairing interstrand crosslinks. Sixteen Fanconi genes have been identified, and when mutated, lead to the development of Fanconi anemia syndrome in humans, a disorder characterized by the loss of

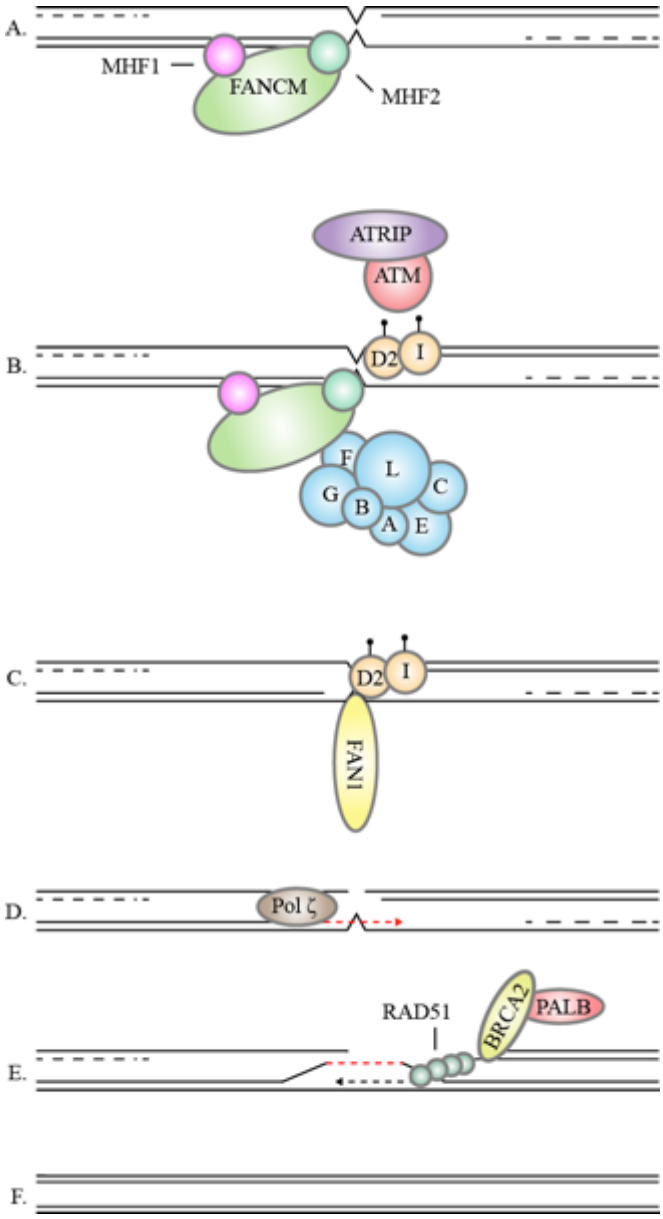


Figure 1.4: Interstrand crosslink repair by the Fanconi anemia pathway. A. The Fanconi anchor complex associates with the DNA lesion**.** B. The Fanconi core complex (Blue) is recruited to the lesion via the anchor complex. The core complex recruits FANCD2/FANCI, which are phosphorylated by the ATRIP/ATM complex. C. The phosphorylated D2/CI complex serves a recognition site for the endonuclease FAN1, which separates the crosslinked strands. D. Translesion synthesis is carried Polζ . E. RAD51 and BRCA2 facilitate strand evasion and H. homologous repair restoring the two dsDNA molecules.

erythroid, myeloid, and lymphoid lineages of blood, and a predisposition to cancer [36].

Insterstrand crosslinks are recognized during S-phase by fanconi anemia complementation group M (FANCM), which associates with centromere protein S (CENPS/MHF1) and centromere protein X (CENPX/MHF2) at sites of damage forming an anchor complex on the DNA [37-39] (Figure 1.4A). The core Fanconi complex then is recruited to the anchor, which consists of fanconi anemia complementation group E, F, G, C, A, L, B, and is responsible for the ubqitination of fanconi anemia complementation group D2 (FANCD2) and fanconi anemia complementation group I (FANCI) and the further recruitment of additional DNA repair factors [40, 41](Figure 1.4B). FANCD2/FANCI complex is phosphorylated by Ataxia telangiectasia mutated (ATM)/ataxia telangiectasia and Rad3-related protein interacting protein (ATRIP), leading to stabilization of the complex to the DNA binding [42] (Figure 1.4B). The association of FANCD2/FANCI to the DNA leads to its ubiquitination by the core complex and recruitment of FANCD2/FANCI-associated nuclease 1 (FAN1), a nuclease that is responsible for releasing the crosslinked strand [36](Figure 1.4C). After strand release, translesion synthesis is carried out by the error prone DNA polymerase ζ at the site of DNA damage [43](Figure 1.4D). The damaged DNA is then repaired by Fanconi specific DNA repair factors breast cancer 2 (BRCA2), breast cancer 1 (BRCA1) interacting protein C-terminal helicase 1 (BRIP1), partner and localizer of BRCA2 (PALB2), and radiation sensitive 51C (RAD51C) as well as factors involved in homologous recombination repair such as BRCA1 and radiation sensitive 51 (RAD51)[36](Figure 1.4E,F).

It has been demonstrated that FANCD2 can disassemble the histone 3 (H3)/histone 4 (H4) tetramers *in vitro,* and FANCD2 depletion leads to alterations in H3 mobility in the genome [44]. Recently it has been reported that the FANCD2 contains a reader domain that can recognize

H4K20 methylation (unpublished results) [45]. It suggests that histone marks could influence ICLR activity by direct interactinons between repair factors and the histone code.

Single strand break repair

A majority of single strand DNA breaks arise from endogenous sources of reactive oxygen species (ROS) and are repaired by the SSBR pathway. The generation of hydrogen peroxide through normal cellular metabolism leads to the formation of free radicals that can act directly on DNA by modifying bases. This type of damage is initially repaired by the BER pathway. As the BER pathway nears completion of the repair, DNA-(apurinic or apyrimidinic site) lyase (APE1) endonuclease initiates the cleavage event that leads to the generation of single stranded DNA (ssDNA) (see Base excision repair section) (Figure 1.5A). The ssDNA is generated as an intermediate product, and is then funneled into the SSBR pathway.

ROS exposure can also cause direct excision of the bases generating a gap and a single strand break in the DNA that is repaired by SSBR. This type of break is recognized by the PARP1 enzyme. PARP1 is a direct sensor of single stranded DNA and binding to ssDNA leads to its enzymatic activation in an allosteric fashion [46])(Figure 1.5B). How PARP1 is able to rapidly sense single strand DNA breaks on a genome-wide scale can be attributed to the amount of the protein present in the nucleus. It is estimated that there is one PARP1 protein for every 20 nucleosomes. [47]. Activated PARP1 catalyzes the formation of poly(ADP-ribose) chains onto H1 and the H3/ H4 tetramers surrounding the DNAdamage site with the aid of histone PARylation factor 1 (HPF1) [48]. X-ray repair cross-complementing protein 1 (XRCC1) is guided to the site of DNA damage via aprataxin-PNK–like factor (APLF) through its PAR-binding zinc finger domain were it associates with H3/H4 tetramer and PARP1 [49, 50]XRCC1

acts as a scaffold for the recruitment of DNA Ligase III and DNA polymerase β, which fill in the gap in the damaged DNA [15, 51](Figure 1.5C).

Lastly, another source of single strand breaks is abnormal Topoisomerase I (TOP1) activity [52]. TOP1 can become permanently linked to DNA at cleavage sites in situations in which free radicals have damaged the DNA [53](Figure 1.5A). Covalently fixed TOP1 is cleaved from the DNA by Tyrosyl-DNA phosphodiesterase-1 (TDP1) [54]. XRCC1 serves as a scaffold between TDP1 and the single strand break repair complex mentioned above [55](Figure 1.5B).

It has been observed that PARP1 PARylation activity occurs at the same sites as other modifications on core histones. Histone 2A lysine 13 (H2AK13), histone 2B lysine 30 (H2BK30), histone 3 lysine 27 (H3K27), histone 3 lysine 37 (H3K37), and histione 4 lysine 17 (H4K17) were identified as sites of PARP1 dependent PARylation [56]. It was found that H4K16 acetylation limits PARylation of H4 histones [56]. PARylation could be limited at sites of single strand DNA breaks by the activity of lysine acetyltransferase 5 (Tip60), which specifically modifies H4K16 at sites of DNA damage [57]. In this scenario, H4 acetylation would dampen the DNA damage signal propagated by PARP1 and would in turn limit the amount poly(ADP-ribose) chains generated. In this way, histone modifications prevent hyper-PARylation, which can be lethal to cells experiencing genotoxic stress [48]. It would be interesting to explore the effects of different histone modifications on single strand break repair.

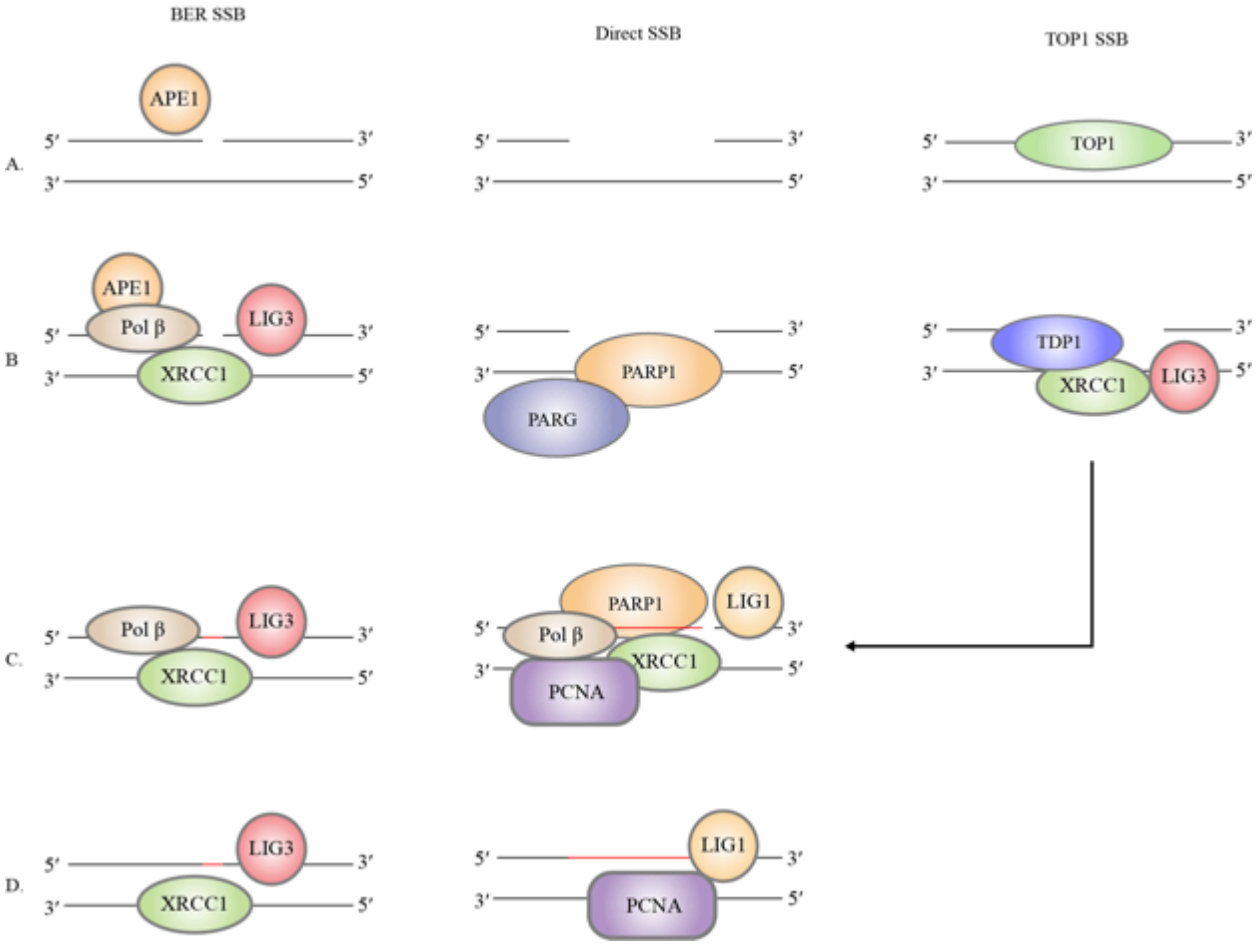


Figure 1.5: Single strand break repair. The activity of BER generates SSB that are processed by the SSBR pathway. A) APE1 introduces a break in one strand following BER. XRCC1, Polβ, and LIG3 are recruited to the DNA break site. C) Polβ fills in the gap and (D) LIG3 seals the nick. Direct SSB (A) can be generated by ROS activity. B) These lesions are recognized PARP1/PARG and (C) recruit PCNA/XRCC1/LIG1 and Polβ fills in the gap and (D) LIG1 seals the nick. TOP1 SSB can be generated by TOP1 becoming permanently linked to DNA through contact with an abnormal base. B) TDP1 cleaves TOP1 from the DNA generating a SSB, which recruits XRCC1 and LIG3. This lesion is then repaired by a similar mechanism as a direct SSB. Modified from [9].

Double strand break repair

DNA double strand breaks are the most rare form genotoxic stress experienced by cells, with fewer than two occurring per cells [4]. However, across species, they represent the most toxic of lesions with 1-2 unrepaired double strand breaks considered lethal [58]. Mammalian studies indicates that 1-10 unrepaired double strand breaks is lethal [59].

Mammalian cells have evolved a complex signaling pathway to sense and repair DNA double strand breaks (Figure 1.6). Kinetics experiments estimate that repair takes place within 3 minutes and up to 720 minutes, depending on the double strand break repair pathway used [60]. The "fast acting" repair pathway identified in early kinetic experiments was hypothesized to depend on DNA ligases; and was later named non homologous end joining (NHEJ). The "slow repair" pathway was dependent on recombinases and later name homologous recombination (HR) [61].

Two independent sensors, PARP1 and X-ray repair cross-complementing protein 6 (XRCC6/KU70)/ X-ray repair cross-complementing protein 5 (XRCC5/KU80) sense double strand breaks and depending on which sensor is utilized, determines which double strand break repair pathway is used. PARP1 directly senses double strand breaks after homodimerization caused by the discontinuity in the broken DNA [62, 63]. The dimerization of two PARP1 molecules on opposite sides of a double strand breaks leads to auto-PARylation and this initiates a cascade of break sensing [62]. PARP1-mediated PARylation leads to the recruitment of single stranded binding homologue 1 (SSB1) and the MRN complex composed of Meiotic Recombination 11 homolog (MRE11)/radiation sensitive 50 (RAD50)/NLR family pyrin domain containing 2 (NLRP2/NBS1), through association of the PAR binding domains of SSB1 and MRE11 [64, 65]. Another study showed that the initial recruitment of MRN to DNA damage primarily depends on SSB1, however the PAR binding domain of MRE11 could fine-tune the localization of MRN [66].

MRN localization to the double strand break recruits ataxia-telangiectasia mutated (ATM), which phosphorylates H2AX on Ser139 (γH2AX) and leads to the recruitment of mediator of DNA damage checkpoint 1 (MDC1) [67]. The recruitment of MDC1 to break

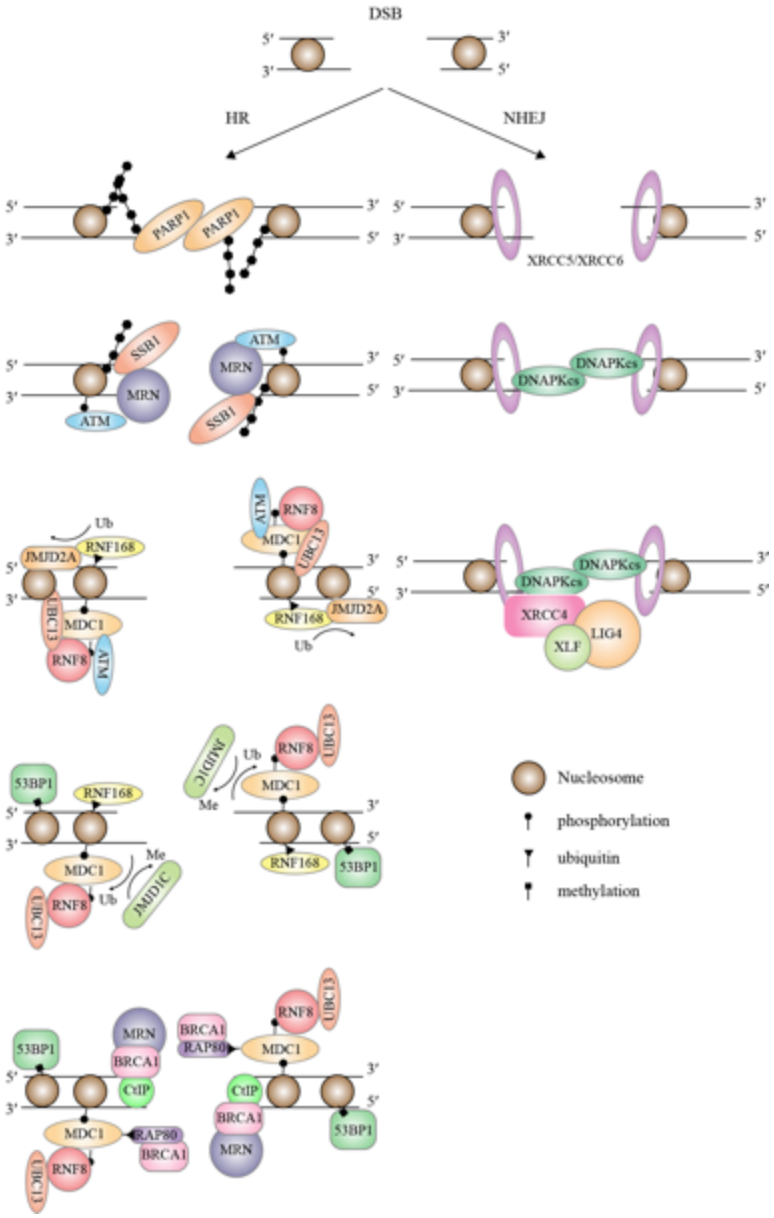


Figure 1.6: Early events in Double strand break repair. When a DSB occurs it can be sensed by PARP1 or by XRCC5/XRCC6. PARP1 recognition of a DSB initiates the HR repair pathway. PARylation of local histones and of PARP1 itself lead to the recruitment of MRN and the subsequent phosphorylation of H2AX. The generation of γH2AX leads to the recruitment of MDC1, which serves as a platform for the DNA repair machinery to assemble. MDC1 loaded ubiquitin proteins work to degrade local proteins expose histone methylation marks necessary for 53BP1 to bind and the extension of ubiquitin chains recruit the RAP80-BRCA1 complex. Both RAP80-BRCA1 and 53BP1 are required for HR to take place. XRCC5/XRCC6 recognition of a DSB initiates the NHEJ repair pathway. XRCC5/XRCC6 recruit DNAPKcs to DSB and these proteins work to hold the two broken ends in close proximity while XRCC4/XLF/LIG4 is recruited to the break site and LIG4 works to seal the broken ends.

sites represents a fork in DNA repair pathway utilization and depending on the phase of the cell cycle, will dictate whether the NHEJ or the HR repair pathway is used for the completion of the repair.

MDC1 directed HR and NHEJ repair begins with the same signaling cascade. HR repair pathway has two sub-pathways: homology directed repair (HDR) and single strand annealing (SSA) [68]. Despite being a part of the HR repair pathway, SSA is always mutagenic. The NHEJ repair pathway also has two sub-pathways: classical NHEJ (cNHEJ) and alternative NHEJ (altNHEJ), both of which are mutagenic [69].

ATM phosphorylation of MDC1 leads to the recruitment of ubiquitin conjugating complex containing ring finger protein 8 (RNF8)/ E2 ubiquitin-conjugating protein 13 (UBC13) complex[70]. RNF8 recruitment stabilizes jumonji domain containing 1C (JMJD1C), a histone demethylase, at sites of DNA damage, which demythylates MDC1 allowing for RNF8/UBC13 to add ubiquitin chains to MDC1 [71]. The ubiquitination of MDC1 then leads to the recruitment of the ubiquitin interaction motif containing 1 (UIMC1/RAP80)/BRCA1 complex [72, 73]. The RAP80-BRCA1 complex fine-tunes HR repair by limiting the length of ssDNA generated by the MRN-BRCA1/retinoblastoma binding protein 8 (RBBP8/CtIP) complex [72].

UBC13, the binding partner of RNF8, extends ubiquitin residues on the linker histone H1, which serves as a loading platform for RNF168 [74]. RNF168 and RNF8 work together to ubiquitinate and degrade jumonji domain containing 2A (JMJD2A), a histone demethylase, which demethylates histone 4 lysine 20 (H4K20) residues [75]. Ring finger protein 168 (RNF168) binding also leads to the ubiquitination of H2A at K12/K15 [74]. The clearance of JMJD2A exposes H3K20me and together with H2A K12/K15 ubiquitination, serves as signal for

tumor protein p53 binding protein 1(TP53BP1/53BP1) to bind [76]. The recruitment of 53BP1 limits end resection carried out by BRCA1-CtIP [77]. In this way, both BRCA1 and 53BP1 fine-tune the amount of ssDNA generated at break sites.

During HR repair, the amount of ssDNA generated following a double strand break will influence whether HDR or SSA takes places. If BRCA1 and 53BP1 are correctly coordinated at a break site, HR will take place. Currently, the absolute length requirements for HR repair are not known. Instead there only exists a nebulous goldilocks idea of length requirements; just the right amount leads to HDR and too much results in utilization of the highly mutagenic SSA repair pathway.

There is no hard-and-fast rule for ssDNA length requirements for HDR and SSA. Instead it is more likely the utilization of HDR or SSA depends on the sequence composition of the ssDNA and binding kinetics of ssDNA binding proteins. Replication protein A (RPA) is the most abundant ssDNA binding protein in the cell and participates in DNA repair [78, 79]. RPA has the preferentially binds ssDNA with polypyrimidine tracts of 8 nucleotides in length, and upon binding, shifts its conformation to occupy up to 30 nucleotides in a 5′ to 3′ direction [80]. The associated of RPA with ssDNA can generate two structures at the sites of DSBs: one condensed structure containing tightly packed RPA coated ssDNA and the other is a linear structure of RPA coated ssDNA, and the linear structure requires the ssDNA to be under constant tension otherwise it will condense [81]. This observation could provide insight into the utilization of HDR over SSA. SSA is mediated by radiation sensitive 52 (RAD52) binding with RPA in a concentration dependent manner to facilitate ligation and annealing of homologous sequences [77, 82].

If the ssDNA generated by end resection is not maintained under sufficient tension, it condenses and the local concentration of RPA increases, driving RAD52 binding and leading to SSA[81]. Evidence suggests that RAD52 is recruited to DNA break sites via switch/sucrose non-fermentable (SWI/SNF) Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4 (SMARCA4/BRG1), the ATPase subunit of the Switch/Sucrose non-fermentable (SWI/SNF) chromosome remolding complex [83]. In this model, RAD52 is sequestered at end of the interface of the double stranded and single stranded region. In order to prevent SSA from occurring, the ssDNA would have to remain elongated, thereby limiting the local concentration of RPA. The elongated RPA coated ssDNA is a substrate for split hand/foot malformation type 1 (SHFM1/DSS1)/BRCA2/RAD51 complex. DSS1 mimics the DNA binding substrate of RPA and causes its dissociation from the ssDNA, allowing RAD51 to be loaded in its place [84]. The association of RAD51 to the ends of the single stranded DNA is sufficient to induce strand invasion and D-loop formation [85]. The formation of the D-loop initiates DNA synthesis and the resulting holiday junction can be resolved in two ways; through synthetic Lethal of unknown (X) function homolog 1 (SLX1)/ synthetic Lethal of unknown (X) function homolog 4 (SLX4) resolvases, or through the Bloom syndrome RecQ like helicase (BLM)/RecQ-mediated genome instability complex (RMI)/ Topoisomerase (DNA) III Alpha (TOPIIIα) pathway [86].

The processes of HDR and SSA are not the only repair pathways that require the formation of ssDNA. The altNHEJ repair pathway, also known as microhomology mediated end joining (MMEJ) requires the generation of ssDNA to function. The mechanics for generation of ssDNA is similar to other HR pathways, however the activity of CtIP is uncoupled from BRCA1. It was found that the unphosphorylated form of CtIP can perform end resection independently of

BRCA1 [87]. The association of CtIP with the MRN complex is sufficient to stimulate end resection activity at sites of DNA damage [88, 89]. The amount of homology necessary for MMEJ/altNHEJ is estimated to between 1 and 4 nucleotides [90]. However, this is not reflective of the actual length of the single stranded DNA generated. It should be noted that the ssDNA generated during this process is a substrate for RPA, which limits MMEJ/altNHEJ activity [91]. The footprint for RPA binding is 8-30 nucleotides, and the ssDNA generated necessary for MMEJ/altNHEJ is at least this long.

How then might RPA be bypassed to facilitate MMEJ/altNHEJ? One possible scenario is that the initial DNA break detection by PARP1 recruits splA/ryanodine receptor domain and SOCS box containing 1 (SSB1), which binds with MRN and stimulates end resection activity prior to BRCA1 recruitment [66]. PARP1 directly stimulates MMEJ via recruitment of XPF/ERCC1/POLθ and XRCC1/LIG3 complexes [92, 93]. The recruitment of these complexes leads to the excision of overhanging ssNDA, gap fill-in, and strand ligation to seal the break.

There exists a continuum of ssDNA that facilitates different types of repair: SSA being on one extreme and cNHEJ on the other extreme. There exist two pathways for cNHEJ activation: the first involves SSB1/MRN brake sensing and the other involves XRCC5/XRCC6 direct sensing of the double strand break. As mentioned above when SSB1/MRN sense a break, it leads to the loading of MDC1 and the recruitment of BRCA1 and 53BP1. Both of these factors fine-tune the amount of ssDNA generated at the break site in order to optimize the utilization of HDR. If end resection is completely blocked, it is hypothesized that 53BP1 can direct DNA repair towards cNHEJ.

How 53BP1 facilities cNHEJ is not clear, and little work has been done to address the mechanism. However, there is link between 53BP1 and its ability to stimulate cNHEJ. 53BP1

interacts with retinoblastoma transcriptional corepressor 1 (RB1), which serves as a scaffold for XRCC5/XRCC6 binding [94, 95]. Therefore, a possible scenario is that the rapid accumulation of 53BP1 at sites of double strand DNA breaks limits the amount of end resection taking place and recruits XRCC5/XRCC6, leading to the activation of the cNHEJ repair pathway.

The XRCC5/XRCC6 heterodimer recruits protein kinase, DNA-activated, catalytic polypeptide (PRKDC/DNA-PKcs) to the sites of DNA damage in order to keep the two broken ends in close proximity. DNA-PKcs localization recruits APE1, tyrosyl-DNA phosphodiesterase 1 (TDP1), polynucleotide kinase 3'-phosphatase (PNKP), which are required for ligation of the broken ends by the XRCC4/XLF/LIG4 complex [96]. The XRCC5/XRCC6 heterodimer is hypothesized to directly sense the double strand DNA breaks by its association with the ends of DNA with high affinity [97] [98]. However, it is not clear on how XRCC5/XRCC6 rapidly detect the double strand break once it occurs. Other DNA damage sensing pathways have components that are in high concentration and activity scanning DNA. One hypothesis is that it is actively recruited to double strand breaks in a PARP1-dependent manner. Mass spectrometry results have identified that XRCC5 was associated with poly(ADP-ribose) molecules [99]. It was found that the PAR domain of XRCC6 is required for cNHEJ repair to take place [100].

This model for PARP1-mediated XRCC5/XRCC6 recruitment to sites of double strand breaks seems at conflict with PARP1 recruitment of HR factors. How can it be that the single activity of PARP1 results in two diverging repair pathways? It was found that XRCC5/XRCC6 heterodimer slides along dsDNA once it has bound to the broken end [98]. It is conceivable that XRCC5/XRCC6 mobility is necessary to make room for the binding of other cNHEJ components. If the histones in front of XRCC5/XRCC6 can be removed, the complex can slide further from the ends of the DNA limiting its ability to recruit other cNHEJ factors and favors

other repair pathways dependent on ssDNA. If histone dynamics are restricted at the break sight, then XRCC5/XRCC6 movement is restricted, thus keeping the complex in close proximity to the break site and favoring cNHEJ activity. However, further test would be needed to determine if XRCC5/XRCC6 can interact with histone remodeling factors necessary for removal of histones.

DOT1L and epigenetic control of DNA damage repair

As mentioned in the above sections epigenetic marks play an important role in DNA damage sensing and DNA damage repair. The mechanism of how specific histone marks contribute to DNA repair processes remains a poorly understood process. In order to shed light on how epigenetic marks facilitate DNA repair we are studying the histone methyltransferase disruptor of telomeric silencing-1-Like (DOT1L). DOT1L is the only known histone 3 lysine 79 methyltransferase in eukaryotes [101-103].

DOT1L was first identified in the budding yeast as "disruptor of telomeric silencing-1" (*DOT1*) in a genetic screen for genes affecting telomere silencing [104]. Biochemical characterization of Dot1 and the mammalian homologue, DOT1L, identified both as methyltransferases responsible for the methylation of H3K79 [101-103]. H3K79 methylation in eukaryotes is broadly associated with euchromatic regions, and is reduced in regions of heterochromatin.

Genome-wide studies of histone marks in *Drosophila* and mammals found that H3K79 methylation occurs within genes undergoing active transcription [105, 106]. DOT1L has experimentally been shown to interact with AF10, and this interaction, in the context of MLL-AF10 leukemic translocation, has been show to play a role in leukemogenesis [107]. Further experiments identified DOT1L as a central component to a transcriptional activation complex,

DotCom, which also contains several common MLL fusion partners AF9, AF10, ENL, and AF17 [108].

Murine hematopoietic cells immortalized with the fusion protein MLL-AF9 show a DOT1L dependence on survival [109]. These observations lead to the development of a small-molecule inhibitor of DOT1L methylation activity that could inhibit the growth of MLL tumors in a xenograft mouse model [110]. It is hypothesized that mislocalization of H3K79 methylation leads to an activation of a leukemia specific transcriptional program [111]. However, evidence from eukaryotes indicates that DOT1L is required for maintaining genome integrity, potentially by its involvement in the DNA damage response [112-116].

In yeast Dot1 knockout caused activation of the pachytene checkpoint in budding yeast [117]. The pachytene checkpoint occurs during meiosis, when non-sister homologous chromosomes pair and undergo recombination. Arrest during this phase is termed the pachytene, or meiotic recombination, checkpoint [118]. In another study using the same *DOT1* deletion in yeast, Δ*dot1,* and two other deletions in genes that prevent meiotic double strand break repair, Δ*dcm1,* and Δ*rad54*, the yeast bypass the meiotic recombination checkpoint and display an accumulation of fragmented chromosomes [119]. These observations suggest that Dot1 plays a role in DSB repair via the HR repair pathway and cell cycle checkpoint control. Indeed, yeast *DOT1* mutants have decreased viability when treated with ionizing radiation (IR) to induce DNA damage; and this phenotype was duplicated in yeast strains carrying mutations at H3K79 that prevent methylation [120]. Altogether, these results indicate that DOT1L and the methylation of H3K79 is important for proper DNA damage response in yeast, and suggests it may be true in mammalian cells.

A specific epigenetic modification system exists in yeast and mammals, called trans-histone modification, in which the deposition of one epigenetic mark facilitates the addition of another. Trans-histone modification was identified in yeast Dot1-dependent H3K79 methylation, which was demonstrated to require H2B123 ubiquitination via Rad6-Bre1 in yeast [121-123]. In budding yeast exposed to IR, it was demonstrated that defects in H2B123 ubiquitination, caused by a mutation in Rad6, lead to defects in H3K79 methylation via Dot1, and a subsequent block in Rad9 activation [114, 124]. Yeast *DOT1* mutants exposed to IR are defective in the G1 checkpoint and fail to induce the phosphorylation of both Rad9 and Rad53 (the yeast orthologue of 53BP1, and homologue of CHK2, respectively) [113]. In this example, the trans-histone modification system is necessary for propagation of the DNA damage signal, and by eliminating one epigenetic mark, the entire signaling cascade is dysfunctional. DOT1L and DOT1L-methylation is central to the trans-histone modification and the DNA damage response.

Further experiments performed with yeast Δ*dot1* mutants provided mechanistic clues regarding the specific DNA damage repair pathways that DOT1L functions in. In the absence of Dot1, uncapped telomeres were subjected to hyper end resection [112]. Hyper end resection is a phenomenon in which the exonucleases that process broken DNA ends are unchecked and generate long sections of ssDNA, which can be problematic because it is a substrate for highly mutagenic DNA repair pathways in mammals such as alt-NHEJ and SSA.
Hyper end resection was observed in non-telomere regions of yeast Δ*dot1* mutants, where double strand DNA breaks were induced using the HO endonuclease [125].

The dependence of Rad9 activation on H3K79 methylation led to work in mammalian cells, in which it was demonstrated that 53BP1 (the mammalian orthologue of Rad9) recognizes H3K79 methylated residues at the sites of DSB [126]. Another report demonstrated that H4K20

methylation by CRB2 is responsible for the recruitment of 53BP1 to sites of DSB [115]. Neither study ruled out the possibility that both marks are used in co-recruitment of 53BP1. Studies in chicken DT40 cells demonstrated that DOT1L deficient cells still form 53BP1 foci upon irradiation, thus indicating that 53BP1 is independently loaded onto DSB [127]. Indeed this appears to be the case, as it was shown that 53BP1 is recruited to sites of DSB by H2AK15 ubiquitination via RNF168 and H4K20 methylation [128]. These observations indicate that DOT1L and H3K79 methylation mediates a DNA damage response independent of 53BP1 activation, and perhaps represents an additional, and novel pathway for DSB repair.

A unique feature of H3K79 methylation is that it can prevent the nucleosome from adopting a compact state [129]. The open state of chromatin facilitated by H3K79 methylation, or DOT1L, itself could allow histone exchange , thereby allowing DNA repair to happen.

The link between DOT1L and histone exchange comes from experiments in mammalian cells implicating DOT1L in complex with factors involved in histone remodeling. First, DOT1L forms a complex with v-myc avian myelocytomatosis viral oncogene homolog (c-Myc) and CREB-binding protein/ E1A binding protein p300 complex (CBP/p300) [130]. We term this complex MyDOT for simplicity of explanation. The CBP/p300 component of MyDOT acetylates several histone residues including H3K14, H3K18, H3K23, H3K27, H3K56, H4K5, H4K8, H4K12, and H4K16, creating marks for the recruitment of downstream factors [131-133]. The DOT1L component of MyDOT is the sole methyltransferase responsible for the methylation of H3K79.

In the absence of DOT1L, c-Myc forms a repressive complex with HDAC1, a histone deacetylase, and the DNA methyltransferase DNMT1 [130]. HDAC1 can remove acetylation marks from a variety of locations including H3K9, H3K14, H4K5, H4K8, H4K12, and H4K16

[134. The residues H4K5, H4K8, H4K12, and H4K16 were also shown in yeast to be important for H2AZ deposition {Altaf, 2010#360]. Furthermore, c-Myc forms a complex with the histone acetyltransferase Tip60 and the histone remolding factor E1A binding protein p400 (p400), which acetylates H4K16 and catalyzes the exchange of H2AZ respectively [57, 135, 136]. Both Tip60 and p400 participate in active histone exchange during DNA damage detection and repair [137]. Finally c-Myc has been shown to directly inhibit cNHEJ by binding to XRCC6 (KU70) in the XRCC5/XRCC6 (KU80/KU70) complex and blocking DNA-PKc association with XRCC6 (KU70) [138]. Therefore, it is conceivable that the repressive c-Myc/HADC1/DNMT1 complex could inhibit H2AZ exchange at sites of double strand breaks, by removing key acetylation marks.

These observations form the basis for the following model for the role of DOT1L in DNA damage repair: 1) the MyDOT complex is rapidly recruited to sits of DNA damage via its interaction with XRCC6 (KU70); 2) concurrently, MDC1 is recruited to the break site with Tip60/p400; 3) together MyDOT and MDC1/Tip60/p400 work to acetylate and methylate histones, which facilitate the exchange of H2AZ into chromatin (Figure 1.7). Using this model, we predict that in the absence of DOT1L, the repressive c-Myc/HDAC1 complex will form, leading to deacetylation of H3 and reducing the efficiency of H2AZ incorporation by

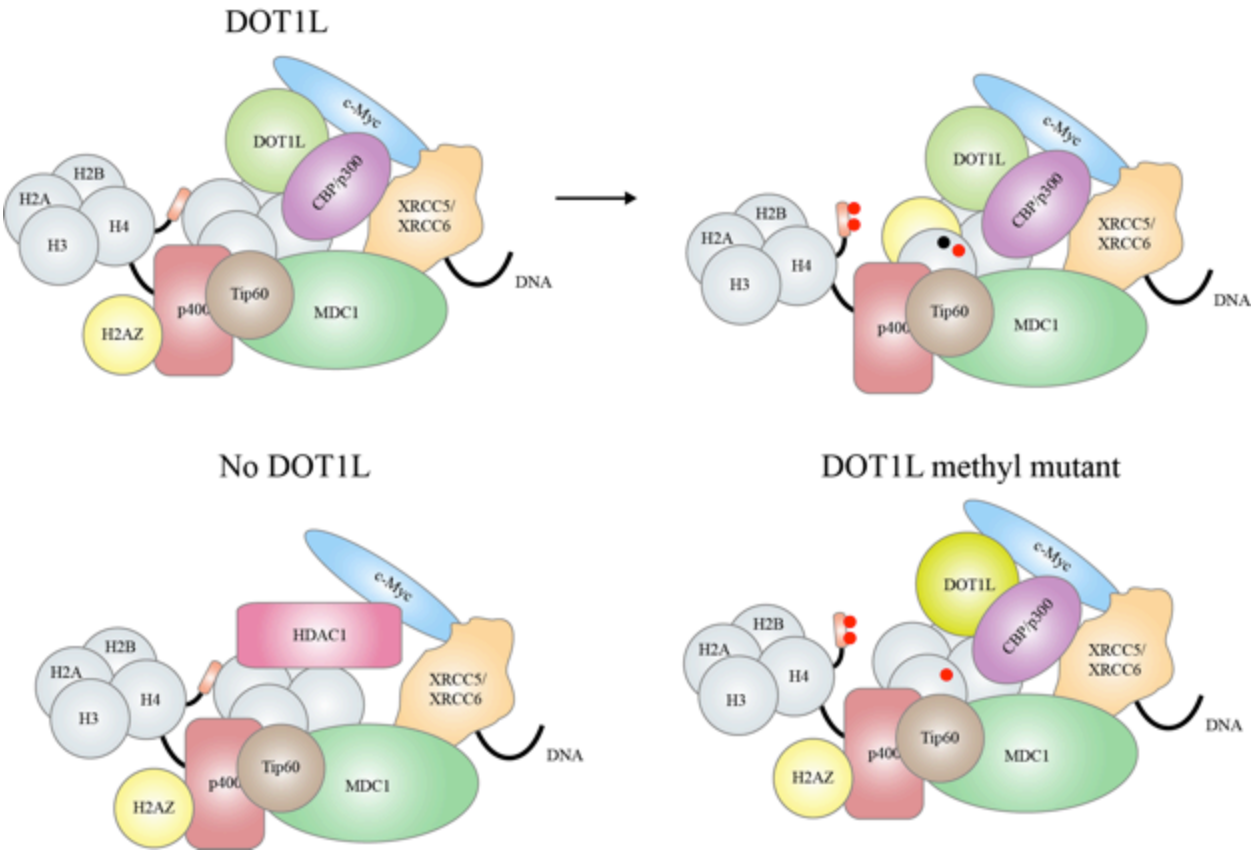


Figure 1.7: Hypothetical MyDOT DNA repair complex. The MyDOT complex consisting of c-Myc, DOT1L, and CBP/p300 is rapidly recruited to sites of DSB via c-Myc interaction XRCC5/XRCC6. MyDOT, together with Tip60/p400 complex methylates (black dots) acetylates (red dots) the local histones to shift them into an open confirmation that is permissive for H2AZ exchange. In the absence DOT1L c-Myc forms a complex with HDAC1, which actively deacetylates histones, which prevents efficient H2AZ exchange. It is hypothesized that a catalytically inactive DOT1L would have a similar effect as a absent DOT1L as the methylation mark could act to block deacetylation activity or facilitate the acetylation.

Tip60/p400. Consequently, an increase in end resection activity will result at DNA damage sites, eventually leading to genomic instability. In the absence of DOT1L H3K79 methylation activity, the MyDot complex will be catalytically inactive, but still facilitate acetylation at the break site. However, the absence of H3K79 methylation could allow for the binding of a deacetylase, such as sirtuin 3 (SIRT3), which has been shown deacetylate H3 marks associated with DNA damage [139]. Indeed, in yeast it was found that H3K79 methylation can activity block the activity of the Sir3 deacetylase [140]. The H3 deacetylation inhibits H2AZ incorporation at sites of DNA damage.

Based on the MyDOT model for DNA damage repair, we hypothesize that in the absence of DOT1L or DOT1L methylation activity, DNA repair would be deficient. Specifically, DOT1L mutants are hypothesized to have defects in H2AZ incorporation, hyper end resection, and defects in DNA repair pathway utilization. Our model provides the potential for novel mechanistic insight in the function of an epigenetic regulator in the DNA repair processes.

To test our hypotheses, we used a combinatorial approach that bridges the fields of biology and computer science. In the following three chapters I will outline three methodologies developed to test the hypotheses. These methodologies include a novel chromatin immunoprecipitation (ChIP) technique that allows for the identification of nucleosomes at DNA double strand break sites in a sequence-independent manner. We also developed a methodology for measuring single stranded DNA, down to nucleotide resolution. Finally, we developed a machine autonomous approach to analyzing microscopy images that incorporates cell cycle data with nuclear foci kinetics. In the fifth chapter I demonstrate the application of these methodologies to elucidate the molecular mechanism of DOT1L function in DNA repair. This approach has led to the discovery of a potential molecular mechanism of DOT1L function in DNA repair. Identification of the molecular mechanism of DNA repair regulated by DOT1L will provide foundation for the improved development of molecular therapy programs designed to induce DNA damage.

Chapter II: Identifying nucleosome dynamics at sites of double strand break

Abstract

Understanding histone dynamics at sites of DNA damage is central to understanding how DNA repair takes place, however, assays for histone dynamics at DNA breaks sites lack the temporal resolution necessary to interrogate a rapid and dynamic phenomenon. Here, we present "broken end ligation chromatin immunoprecipitation" (BEL-ChIP), a method that combines *in situ* hybridization with chromatin immunoprecipitation. BEL-ChIP allows for the identification of histones, histone variants, and epigenetic marks that are present at sites of DNA double strand breaks independent of the underlying DNA sequence. BEL-ChIP will allow for unprecedented temporal resolution of protein composition and modification at sites of double strand breaks.

Introduction

Histone variants and epigenetic marks help guide DNA repair machinery within cells. Chromosome immunoprecipitation (ChIP) is the main technique to map proteins and epigenetic marks to the genome; however, ChIP cannot be used to identify random double strand breaks. ChIP in combination with zinc finger nucleases have been used to map chromosome dynamics at specific break sites [136]. However, using ZFN or CRISPR/Cas9 systems are temporally limiting with the interrogation of break sites usually occurring > 24 hours after transfection, and a prior knowledge of the sequence surrounding the break site is required for qPCR. To map double strand breaks, independent of sequence composition, Direct *in situ* breaks labeling, enrichment on streptavidin and next-generation sequencing (BLESS) was developed [141]. BLESS allows for the mapping of double strand breaks with high resolution using a broken end capture technique combined with next generation sequencing. Despite this advancement for identifying locations of double strand breaks in a sequence independent manner with high temporal

resolution, there is missing a comparable technique for protein composition and modification. Therefore, an improved method for identifying protein composition and modifications with higher temporal resolution is needed.

To address these problems, we developed BEL-ChIP, a method using a modified version of the in situ labeling of BLESS, and the temporal control of double strand break induction using ionizing radiation (Figure 2.1a). In BEL-ChIP, labeling of double strand breaks occurs in situ in the nucleus before lysis, minimizing possible labeling of new double strand breaks introduced through sample processing. ChIP is then performed on the linker labeled library, directly capturing the monosome that contains protein or epigenetic mark of interest. qPCR is applied to the enriched library using adaptors that are specific for the broken end linker. The BEL-ChIP protocol is robust, reproducible, and can be completed in a few days (Figure 2.1a).

Material And Methods

Cell Lines

Wild type HEK293T cells were purchased from ATCC. Mutant HEK293T *DOT1L*^{STOP} and *DOT1L*^{Y312A} are described elsewhere (Chapter IV).

Cell Culture

Wild type HEK293T and mutant HEK293T *DOT1L*^{STOP} and *DOT1L*^{Y312A} cell lines were maintained in 10% FBS, DMEM, Pen/Strep in 37 ^{o}C, 5% CO_2.

Primers and adapters

Broken End Linker	TACTACCTCGAGTAGGGATAACAGGGTAATTTTTTATTACCC TGTTTATCCCTACTCGAGGTAGTA
Anchor Linker	CGTCGTCTCGAGTAGGGATAACAGGGTAATTTTTTATTACCC TGTTTATCCCTACTCGAGACGACG
BEL Primer	TTATCCCTATCTCGAGGTAGTA
AL Primer	TTATCCCTACTCGAGACGACG

Oligonucleotide annealing

The Broken end linker and Anchor linker (IDT) were annealed using T4 DNA Ligase Reaction Buffer (#B0202S, New England Biolabs) to a final concentration of 10μM. The reaction was heated to 98^{o}C for 5 minutes and allowed to cool to room temperature at 1^{o}C per minute.

BEL-ChIP nuclear extract preparation steps

Wild type HEK293T and mutant HEK293T *DOT1L*^{STOP} and *DOT1L*^{Y312A} cell lines were plated into 10 cm dishes two days prior to treatment of ionizing radiation. On the day of experiment, cells were trypsinized, neutralized with 500 μL 10% FBS, 1X DMEM, and centrifuged at 300xg for 7 minutes to pellet. Cells were treated with 10 Gy ionizing radiation following manufactures recommendations. Cells were gently resuspended and placed in a 37 ^{o}C, 5% CO_2 incubator for 60 minutes. Fifty four μL 37% Formaldehyde (2% Final conc.) was added to the sample, and incubated with constant rotation at room temperature for 30 minutes. 62.5 μL of 2 M Glycine (125 mM) was added to the samples and incubated with constant rotation at room temperature for 5 min. Cells were pelleted at 300xg for 7 minutes at 4 ^{o}C and then washed three times with ice cold PBS (-Ca^{2+},Mg^{2+}). Cells were resuspended in 1mL of ice cold Buffer A + 0.5 μL 1M DTT + 5 μL 200X PIC (SimpleChIP Kit #9003, Cell Signaling Technology) and incubated on ice 10 minutes with mixing every 3 minutes. Cells were harvested at 3000 rpm for 5 minutes at 4 ^{o}C. The pellet was washed 3X with 500 μL NEBuffer 2.1 (#B7202S, New England Biolabs). The pellet was resuspended in 88 μL of NEBuffer 2.1, 10 μL dNTP (1mM, #N0446S, New England Biolabs), 2 μL T4 DNA Polymerase (#M0203S,New England Biolabs). The reaction was incubated at RT for 45 minutes with gentle mixing every 5 minutes. The sample was then washed 3X with 500 μL ice cold T4 DNA Ligase Reaction + 0.1% Triton X100

(#T8787, Sigma Aldrich). The pellet is then washed once with 500 µL T4 DNA Ligase Reaction Buffer. The pellet was resuspended in 18.5 µL T4 DNA Ligase Reaction Buffer with 5 µL of 10µM annealed Broken end linker and 1.5 µL T4 DNA Ligase (#M0202S, New England Biolabs). The reaction was incubated 18-20 hours at 16 ^{o}C with constant mixing. The samples were washed 3X with T4 DNA Ligase Reaction + 0.1% Triton X100. The pellets were resuspended in 100 µL ChIP Buffer (SimpleChIP Kit #9003, Cell Signaling Technology), 2 µL RNAse A (10mg/mL, # EN0531, ThermoFisher Scientific) and sonicated 30 seconds on and 30 seconds off for 45 minutes at 4^{o}C (sonicator brand). Samples were centrifuged at 10,000 rpm for 10 minutes at 4 ^{o}C, transferred to new tubes and frozen at -80 ^{o}C.

Chromatin immunoprecipitation (ChIP)

ChIP was performed using an amount of BEL-ChIP nuclear extract corresponding to 3.4 µg of input genomic DNA in 250 µL of ChIP Buffer + 1X PIC (SimpleChIP Kit #9003, Cell Signaling Technology). Antibody was added to manufacture specifications and incubated overnight with end over end mixing at 4 ^{o}C. 30 µL magnetic protein G beads (SimpleChIP Kit #9003, Cell Signaling Technology) were added to the samples and incubated at 4 ^{o}C with end over end rotation for 4 hours. A magnetic rack was used to collect the beads and the sample was washed 2X with 500 µL NEBuffer 2.1. The beads were resuspended in 100 µL of NEBuffer 2.1, 10 µL 1mM dNTP, 2 µL T4 DNA Polymerase and incubated at room temperature for 25 minutes with constant shaking. The beads were placed in a magnetic rack to clear the supernatant and washed 2X with 500 µL ice cold T4 DNA Ligase Reaction Buffer. The beads were resuspended 50 µL T4 DNA Ligase Reaction Buffer with 5µL of 10µM annealed Anchor linker and 1.5 µL T4 DNA Ligase. The sample was incubated overnight at room temperature with constant

shaking. The beads were washed and DNA was purified following the SimpleChIP Kit procedure.

I-SceI Digest

Purified DNA from the ChiP procedure (25 µL) was digested with 1 uL of I-SceI endonuclease (#R0694S, New England Biolabs) for 1 hour and heat inactivated.

PCR

PCR of 10 cycles and annealing temperature 61^{o}C was performed using Q5 High-Fidelity DNA Polymerase (#M0491S, New England Biolabs) and BEL and AL primers on 5µL of I-SceI digested DNA purified from the ChIP procedure.

Gel Purification

PCR reactions were run on a 2% agarose gel (NuSieve™ GTG™, Lonza) and the 200-400 bp region was excised and purified using a QIAquick Gel Extraction Kit (#28704, Qiagen).

qPCR.

qPCR was performed using Power SYBR® Green PCR Master Mix (#4367659, Thermofisher Scientific), BEL and AL primers with annealing temperature of 61 ^{o}C and 40 cycles on ABI 7900 system. Two-tailed Student's t-test was used for significance testing.

Antibodies

H2AZ (#2718, CellSignaling), H4K acetylation (#ab7311, Abcam).

Results

We were interested in understanding the histone and histone modification kinetics at sites of double strand breaks. We performed BEL-ChIP on H2AZ and H4K acetylation in wild type HEK293T cells, as well as two mutant HEK293T cell lines *DOT1L*^{STOP} and *DOT1L*^{Y312A} 60 minutes post ionizing radiation treatment to test our method. Similar to *in situ* ligation (ISL) for

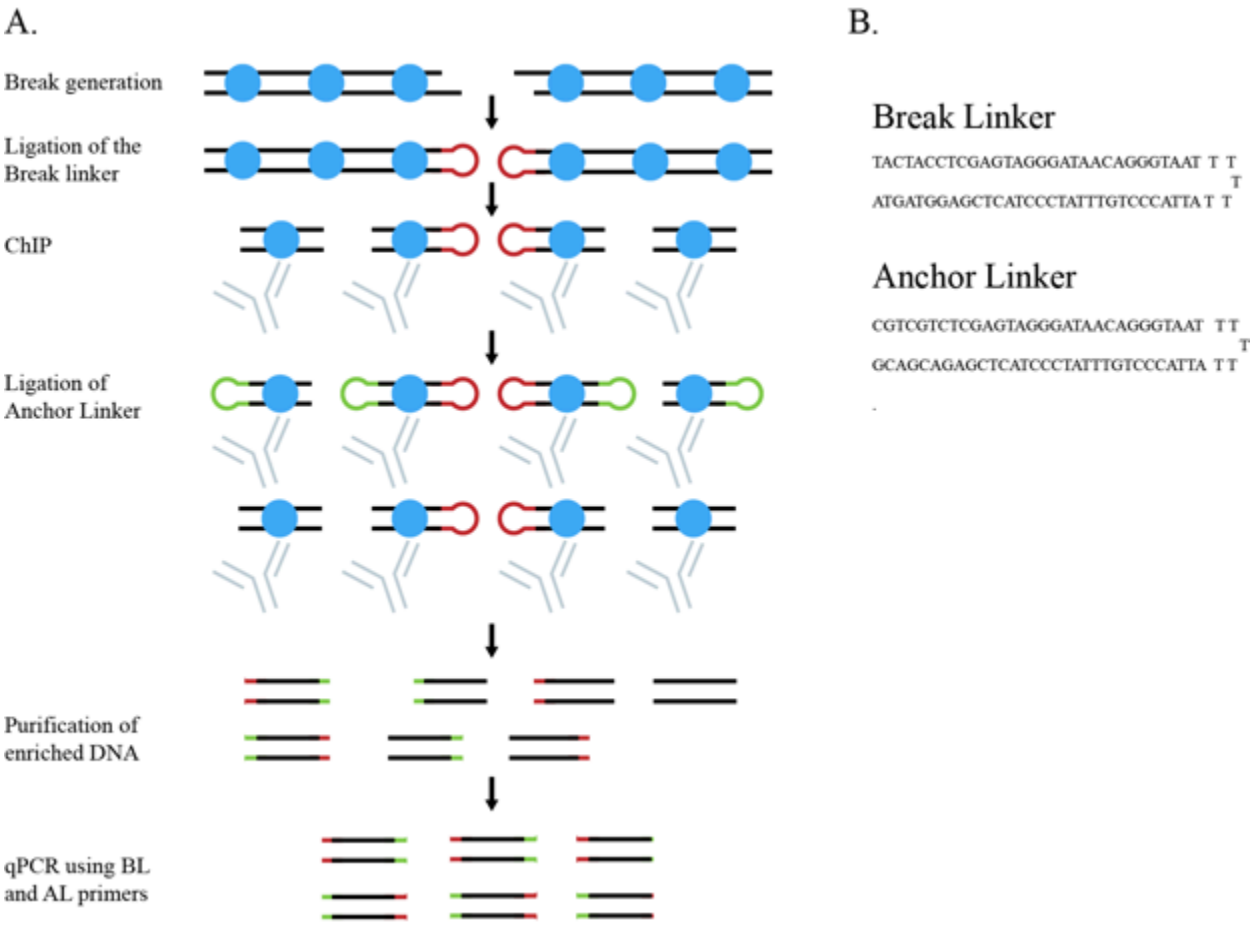


Figure 2.1: BEL-ChIP overview and linker design. A. BEL-ChIP begins with the generation of DSB using ionizing radiation. Cell are harvested and an *in situ* ligation is performed using the Break linker (B) to label the broken end of DNA. Sample is then processed as a normal ChIP procedure. After IP there is an on column ligation step that uses the Anchor linker in (B). The addition of Anchor linker allows for interrogation of enrichment of factors at break sites without having to know were the break occurred.

labeling double strand breaks occurring during apoptosis and BLESS, our method relies on direct *in situ* labeling of induced double strand breaks. However, the design of the hairpin linkers was modified based on knowledge acquired from *in situ* ligation experiments designed to label apoptotic cells [142]. It was found that loopless hairpins have minimal to no background staining, however these require a special modification to the ends to hairpins and for our experiment the modifications are not needed [143]. Therefore, we initially chose a hairpin loop

size of 5 nucleotides, because we reasoned this was reasonable tradeoff between longer stem loops that have higher nonspecific background, and smaller stem loops that might introduce steric hindrance in hairpin formation (Figure 2.1B). A possible area of improvement would be to determine the minimum loop size needed for linker adaptors. This would be especially important as the sensitivity of the assay is increased from qPCR to deep sequencing.

In ChIP protocols it is standard practice to generate a distribution nucleosomes that spread from monosome, disome, trisome and beyond. Since we were interested in the nucleosome composition immediately adjacent to the break site, it was necessary to generate monosomes. Enzymatic treatment of formaldehyde crosslinked nuclei with micrococcal nuclease is a common method to fragment genomic DNA between nucleosomes. However, using micrococcal nuclease would result in the cleavage of the broken end linker. Therefore we settled on a sonication method that robustly generated a large monosome population (Figure 2.2A). We tested whether the monosome population retained the broken end linker by ligation of a second linker to monosome followed by PCR. We were able to demonstrate that the sonication technique did not remove the broken end linker, as we were able to use PCR to amplify labeled monosomes (Figure 2.2B).

After generation of the monosome population using sonication, labeled nucleosomes were enriched using immunoprecipitation. Immunoprecipitation can be performed using antibodies against core histones or the epigenetic marks associated with them. We performed immunoprecipitation of the histone variant H2AZ and the H4K acetylation because they both play a role in double strand break repair [136, 144]. Immunoprecipitation of H2AZ, H4K acetylation and IgG was performed on broken end ligated

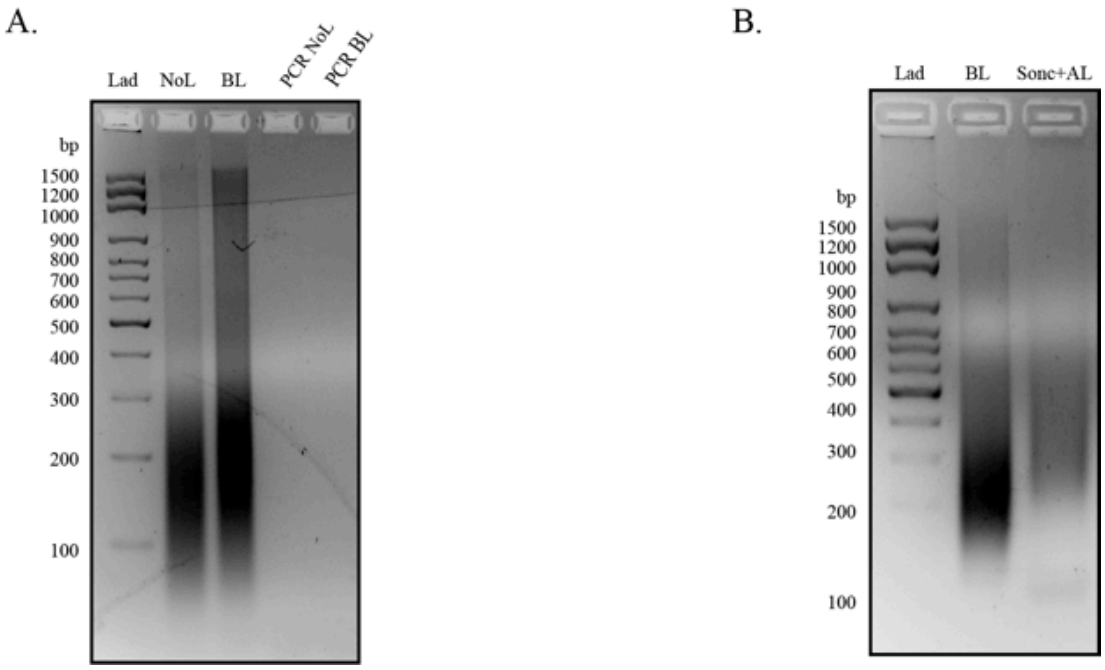


Figure 2.2: BEL-ChIP sonication and linker stability test. A. Lad, DNA loading ladder; NoL, monosome generation from sonication; BL, monosome generation of from sample with the BL linker ligated; PCR NoL and PCR BL, PCR using primers specific for BL. This demonstrates that a single linker ligation is insufficient for amplification. B. BL sonication stability test. After sonication 10 ng of total BL labeled genomic DNA was ligated with the AL and PCR amplified.

genomic DNA extract using a standard ChIP protocol with modifications (Methods). Prior to elution of DNA from the bead:antibody:target complex, a second adaptor was ligated to the enriched DNA fragments. The second hairpin loop linker adds a known sequence to allow for PCR amplification without knowing the underlying genomic sequence.

After purification of the immunoprecipitation, we removed the hairpins using the I-SceI endonuclease and PCR amplified with 10 cycles of amplification. In order to enrich for products corresponding to monosomes, we gel purified the PCR products between 200-400 base pairs. We performed qPCR and found the differential incorporation of H2AZ at sites of double strand breaks between our wild-type and mutant cell lines (Chapter V, Figure 5.17A). H4K acetylation did not show a difference between samples (Chapter V, Figure 5.17A).

Discussion

We have applied BEL-ChIP to study the dynamics of H2AZ and H4K acetylation at sites of double strand breaks. We envision that BEL-ChIP can be used to study histone and histone modification kinetics following the induction of double strand breaks. Coupling BEL-ChIP with ionizing radiation allows for more precise control of the timing of double strand breaks. This offers a significant improvement over the recent uses of CRISPR/Cas9 and zinc finger nuclease induction of breaks because the timing of the break induction cannot be completely known. Though qPCR is used to quantify the dynamics at breaks site, the application of next generation sequencing would allow for the genome wide view of histone exchange at sites of DNA double strand breaks. It is now possible to map the exchange of all histones and their modifications for which there are antibodies using BEL-ChIP. This methodology will give an unprecedented view of the topology of histones and epigenetic marks involved in DNA repair.

Chapter III: Measuring DNA end resection using Oxford Nanopore sequencing

Abstract

Nanopore sequencing developed by Oxford Nanopore Technologies allows for the rapid sequencing of long DNA molecules of heterogeneous lengths. During the process of DNA repair, cells will actively recess DNA ends surrounding the DNA break sites in a process termed "end resection". End resection is necessary for the process of homologous repair to take place, and alterations in end resection can have drastic effects on the type of DNA repair pathway utilized. Hyper and hypo end resection can drive highly mutagenic repair pathways. The current methodologies have genome-wide capabilities, but lack resolution, or have high resolution but require prior knowledge of break location. We developed a novel methodology that allows for nucleotide level resolution of end resection in genome-wide manner.

Introduction

An ideal approach for measuring the kinetics of end resection following DNA damage is to allow for the selection of single stranded DNA molecules from a pool of double stranded DNA with minimal processing. The current methodologies used for measuring end resection are complex, but not ideal, and are limited by requiring either prior knowledge of DNA break sites (as with qPCR assays) or with zero nucleotide resolution (as with DNA stretching) [145, 146]. Furthermore, it is difficult to control the timing of DNA break induction; such is the case with over expression of endonucleases or treatment with genotoxic drugs. An ideal situation is the use of ionizing radiation, however, the location of the double strand breaks is unknown. Genome-wide sequencing could be used to identify hot spots of mutation after ionizing radiation and these regions could be used for qPCR primer design. However, this approach is complex, laborious, and prohibitively expensive.

An improved method would facilitate sequencing of enriched single strand DNA molecules from cells in which the DNA damage could be precisely controlled. Unlike the previous methods mentioned, applying Nanopore sequencing to this problem would allow for sequencing long regions of single stranded DNA generated after induction of double strand breaks, without knowing the sequence composition beforehand.

To generate single stranded DNA that is a product of end resection, cells are exposed to ionizing radiation. This allows for fine control over the timing of DNA damage; far better than can be achieved with over expression of endonucleases and genotoxic drugs. The enrichment of single stranded DNA from double stranded DNA introduces a significant barrier. Solid phase reversible immobilization methods (SPRI) are capable of purifying single stranded DNA, however, only from a starting population of pure single stranded DNA. This is because the same chemistry that drives binding of double stranded DNA works on single stranded DNA. Furthermore, the single stranded DNA generated from end resection is not a free molecule, rather it is covalently linked to the end of double stranded DNA. To that end we developed a method that liberates the single stranded region from the double stranded region while eliminating the double stranded DNA, thus yielding a pool of enriched single stranded DNA. The single stranded DNA is not by itself amenable for analysis and therefore we developed a method for converting single stranded DNA into double stranded DNA in a sequence independent manner.

Because the location of breaks cannot be precisely known when using ionizing radiation, this eliminates the utility of qPCR based approaches. Instead, a method that allows for the measurement of single stranded DNA in sequence independent manner is required. Next-generation sequencing platforms such as Illumina could allow for the identification of genomic regions contained within the single stranded DNA. However, the measurements generated from

an Illumina platform require a genome to align against in order to calculate length. This is a severe limitation, as it has an assumption that the gene model being used is a true gene model. A better methodology would be to measure the length of a single stranded region in a genome-independent manner. Cloning of the single stranded region into a restriction enzyme independent vector such as TOPO (Invitrogen) coupled with Sanger sequencing would provide a genome-independent measurement of length. However, this approach has several severe limitations. First, the length of single stranded DNA would have to be under the maximum sequencing length of the Sanger method (<1000 nt). The cloning method employed has a selection pressure on the size of the molecules inserted into the vector; this could severely skew the measurement of length that is independent of any experiment designed to test end resection rates. Lastly, the cost to sequence individual molecules of single stranded DNA using Sanger is significantly higher than using Illumina platforms.

Current widely adopted sequencing technologies are not suitable for measuring the length of single stranded DNA generated from end resection. Therefore, we applied Nanopore sequencing technology to overcome this problem. Nanopore technology works by passing a nucleic acid molecule through a pore anchored in a membrane. By measuring the voltage change as the nucleic acid transients the pore, the bases of the molecule can be determined. The benefit of this technology for measuring single stranded DNA is that it can measure the full length of a nucleic acid molecule in a genome-independent manner and the read lengths are incredibly long, in some cases exceeding 100 kilobases [147]. Utilizing Nanopore technology allows for length measurement of individual nucleic acid molecules at nucleotide resolution in a genome independent manner.

Materials and Methods

Cell Lines

Wild type HEK293T cells were provided by ATCC. Mutant HEK293T *DOT1L*^{STOP} and *DOT1L*^{Y312A} are described else were (Chapter IV).

Cell Culture

Wild type HEK293T and mutant HEK293T *DOT1L*^{STOP} and *DOT1L*^{Y312A} cell lines were maintained in 10% FBS, DMEM, Pen/Strep in 37 ^{o}C, 5% CO_2.

Enrichment of ssDNA

Wild type HEK293T and mutant HEK293T *DOT1L*^{STOP} and *DOT1L*^{Y312A} cell lines were exposed to 10 Gy of ionizing radiation and total genomic DNA was isolated at 60 and 240 min post radiation exposure. Genomic DNA was isolated by adding resuspending cell pellets in 50 µL of Extraction buffer (Tris-HCl, pH 9.5, 0.1M, KCl 0.25M,EDTA 0.01M), 10 µL Proteinase K (20mg/mL), 10 µL RNAseA (10mg/mL). The samples were incubated at 55^{O}C for 15 minutes and ethanol precipitated. The extracted DNA was digested with RNAseA and dsDNA overnight at room temperature. The samples were ethanol precipitated and twice.

Coversion of ssDNA to dsDNA

ssDNA was converted to dsDNA using the following reaction 5 µL dNTP, 2µL AMPLigase, 2µL T4 DNA polymerase, 10µL Random hexamer, 5 µL Ligase buffer. Reaction was incubated at 25^{o}C for 10 min and 16^{o}C for 16 hrs.

Results

In our approach to measuring single strand DNA following induction of DNA damage, we first developed a method to isolate single stranded DNA from cells that have been treated with ionizing radiation. We first developed a method that would extract DNA from cells with as little to no mechanical manipulation. We used a high salt/Protease K/RNaseA method for

extraction of total DNA. The benefit of this method over others is that it allows for near total removal of proteins and RNA from the sample in a short amount of time. The remaining nucleic acids were precipitated.

To enrich single stranded DNA, it is necessary to differentiate between true single stranded DNA and contaminations of genomic DNA in a reverse transcriptase reaction. To accomplish this, we employed the use of a genetically engineered DNase, dsDNase (Thermoscientific), which specifically and robustly digests double stranded DNA. We used dsDNase and RNase to further enrich for single stranded DNA.

Currently, single stranded DNA alone is not amenable for sequencing using Nanopore technology and requires double stranded DNA for both physical tethering to the pore complex, and sequencing both strands of double stranded DNA results in fewer sequencing errors. Therefore we developed methods to convert single stranded DNA to double stranded DNA in a non-amplifying manner. The first problem we encountered was the generation of a complement strand without prior sequence knowledge. We utilized methodologies used for reverse transcription, and employed random hexamer primers to prime the single stranded DNA. The use of random hexamers allows for the priming of regions whose sequence is not known. Typically, random priming is coupled with a strand displacing polymerases. Polymerases with strand displacement activity allow for the amplification of low abundant molecules in RNA and DNA populations. However, an undesired effect of strand displacing polymerases is the amplification of specific regions over others. This is a result of the increased frequency of the occurrence of some hexamers over others in the genome. We desired to only generate one complete complement strand per single stranded DNA molecule. Therefore, we employed a non-strand displacing DNA polymerase, T4 DNA polymerase (New England Biolabs) in conjunction with a

DNA ligase used for second strand cDNA synthesis. In this way complement strand generation can be conceptualized as a fill-in method. The final step in our library preparation is to ligate hairpin adaptors to the each end of the molecules, which allows for the amplification library.

We applied this methodology to analyze end resection in wild-type, $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutant HEK293T cell lines. The yeast *Δdot1* mutant has a hyper end resection phenotype, however, this has not been observed in mammals. We exposed wild-type, $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutant HEK293T cell lines to 10 Gy of radiation and isolated nucleic acids following 60 minutes and 240 minutes (see Methods details). We converted the single stranded DNA into a double stranded sequencing library and PCR amplified the library (Chapter V, Figure 5.18). We followed the manufacture specifications for library preparation for Nanopore sequencing and barcoding. We barcoded each sample (nine in total) and combined them to generate a single library for sequencing.

Sequencing was carried out as described in the manufacture's protocol for Nanopore sequencing. Metrichor (Metrichor.com) was used to call bases from the MinION sequencer. MinION sequencing output decreases with time because contact with the nanopores is driven by diffusion (Figure 3.1a). We added new sequencing library at 7.5 and 12.5 hours, as indicated by the uptick in yield at those time points (Figure 3.1a). The total yield for the sequencing run was 73.78 million bases from approximately 19,582 individual molecules (Figure 3.1b). The average read length was 2.17 kb with the longest read to be 70.46kb. Quality scores for Nanopore reads are given in qscores with values from 0-14; 0 being the worst and 14 the best. The average qscore for the sequencing run was 9.7, which is above the cut off of 9. The quality of the reads did not decrease with time and remained the same through the course of the sequencing run (Figure 3.2a). A majority of the sequenced DNA was below four kilobases, and sequencing reads

below a qvalue of 9 were discarded (Figure 3.2b). Sequencing reads were further processed to identify the ones that only contained barcodes (Figure 3.3).

We compared the length of single stranded DNA from wild-type, *DOT1L*^{Stop} and *DOT1L*^{Y312A} mutant cells at different time points. We identified sequences that were common in wild-type and *DOT1L*^{Stop} mutant samples at each time point and plotted their distributions (Figure 3.4a). We performed the same comparison on the wild-type and *DOT1L*^{Y312A} mutant (Figure 3.4b). The overall distribution of lengths between the mutants and wild type are similar (Figure 3.4). However, *DOT1L*^{Stop} and to a lesser extent *DOT1L*^{Y312A} mutants have more sequences in the range of 2700-3200 bases than wildtype at 60 and 240 minutes post ionizing radiation (Figure 3.4). *DOT1L*^{Y312A} mutants also have more sequences 3500 bases, indicating this mutant can generate longer single strand DNA than wild-type after DNA damage (Figure 3.4).

These observations are similar to what we observed globally using a PCR assay (Figure 5.18, Chapter V), however, the PCR assay lacks resolution of sequencing. We wanted to test the hypothesis that a lack of DOT1L and its methylation activity will generate longer single stranded DNA after DNA damage. We compared the mean differences in lengths between regions of single stranded DNA that were common between the wild-type and *DOT1L*^{STOP} mutant, and the wild-type and *DOT1L*^{Y312A} mutant, (Figure 3.5,3.6).

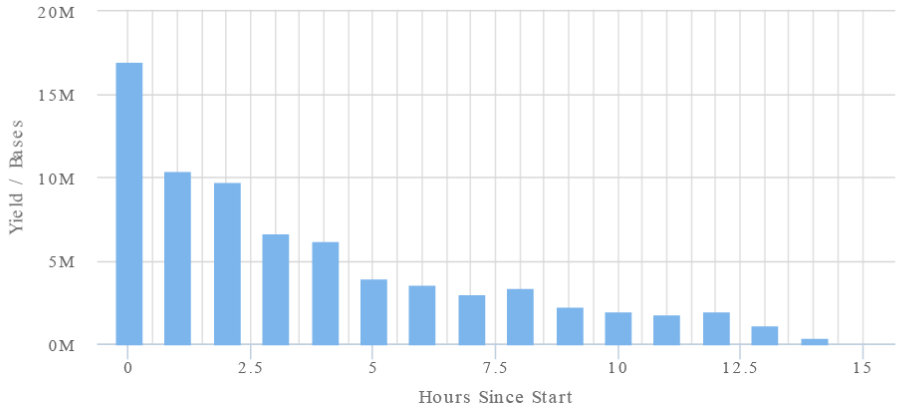
We analyzed all the common sequences between wild-type and *DOT1L*^{STOP} mutant and found that *DOT1L*^{Stop} mutant has more single stranded DNA than wild-type at 60 and 240 minutes post radiation treatment (Figure 3.5a vs. b). A similar analysis was performed on common sequences between wild-type and *DOT1L*^{Y312A} mutant (Figure 3.6). We found that rates of ssDNA generation are similar between the two, but the *DOT1L*^{Y312A} mutant generations longer ssDNA at 60 minutes post irradiation (Figure 3.6a vs b).

It is of note that at 60 minutes post irradiation we identified that *DOT1L*^{Stop} and *DOT1L*^{Y312A} mutant cells have defects in H2AZ incorporation at sites of DNA damage (Chapter V, Figure 5.17). H2AZ is a factor that has been shown to limit end resection activity [136]. Our observations of H2AZ incorporation defects coupled with identification of ssDNA generation defects provide a link between histone variant exchange and end resection activity.

Discussion

We have applied Oxford Nanopore sequencing technology to study the kinetics of end resection following the induction of double strand breaks. We envision that this methodology will allow for more accurate measurement of end resection kinetics. The nucleotide resolution of sequencing allows higher resolution measurements on a genome wide scale, which is currently not possible with current methods for measuring end resection activity

A.



B.

	Total	Filtered
Total Yield	73.78 M Bases	73.78 M Bases
Sequence Length - Average	2.17 K bases	2.17 K bases
Sequence Length - Median	2.31 K bases	2.31 K bases
Sequence Length - Mode	3.22 K bases	3.22 K bases
Longest Read	70.46 K bases	70.46 K bases
QScore - Average	9.7	9.7
QScore - Median	9.4	9.4
QScore - Mode	7.8	7.8

Figure 3.1: Oxford Nanopore sequencing report. A. Basepair yield over the course of the sequencing run. B. Summary table for sequencing metrics for the run.

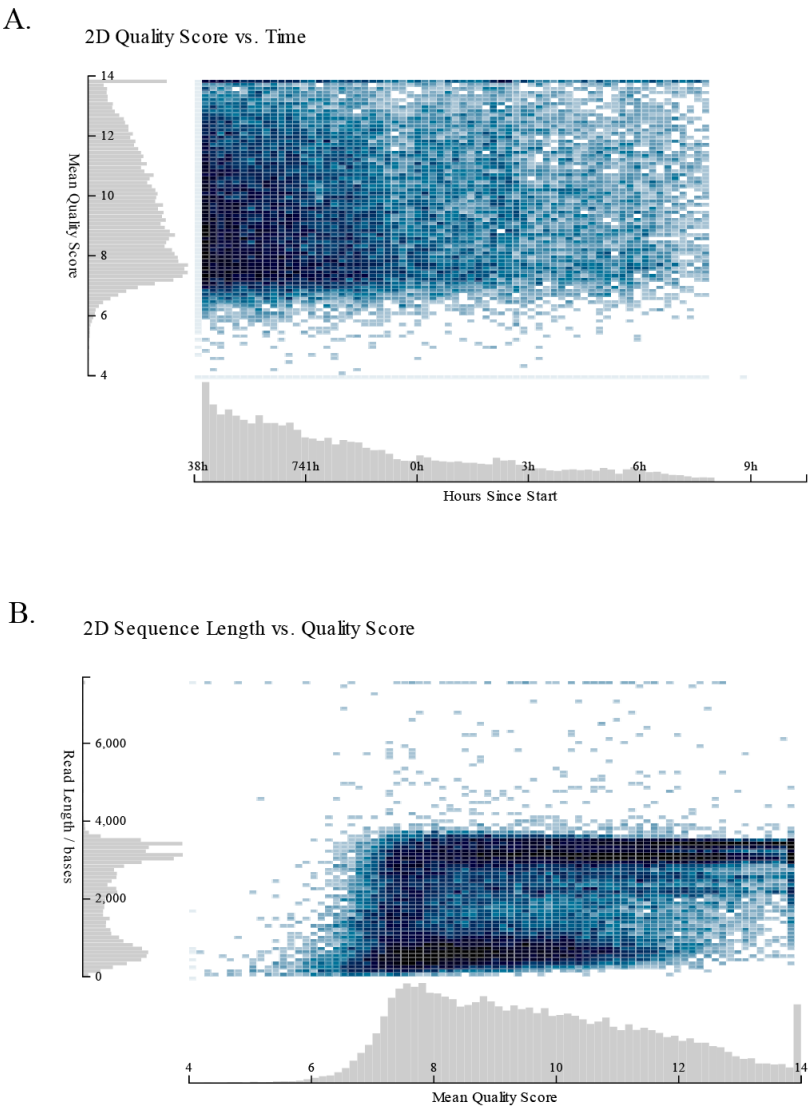


Figure 3.2: Oxford Nanopore sequence length vs quality report. A. 2D sequencing data quality score vs. time graph. The x-axis is the time since start of experiment and the y-axis is the mean quality score (qscore). B. 2D sequence length vs quality score. The x-axis is the mean quality score and the y-axis is the read length in bases.

A.

Barcode ID	Read Count
BC01	804
BC02	863
BC03	594
BC04	1971
BC05	1130
BC06	1370
BC07	1168
BC08	1172
BC09	290

B.

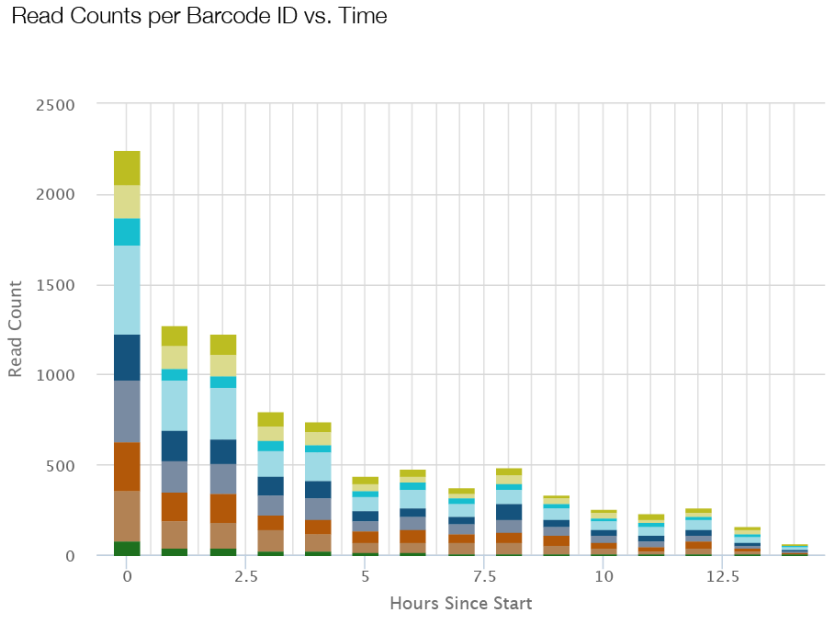


Figure 3.3. Oxford Nanopore barcode recovery report. A. Number sequences recovered for each barcode. BC1-BC3 wild-type samples no irradiation control, 60 and 240 minutes post ionizing radiation. BC4-BC6 *DOT1L*^{STOP} samples no irradiation control, 60 and 240 minutes post ionizing radiation. BC7-BC9 *DOT1L*^{Y312A} samples no irradiation control, 60 and 240 minutes post ionizing radiation. B. Read count output for each barcode over the time of the experiment.

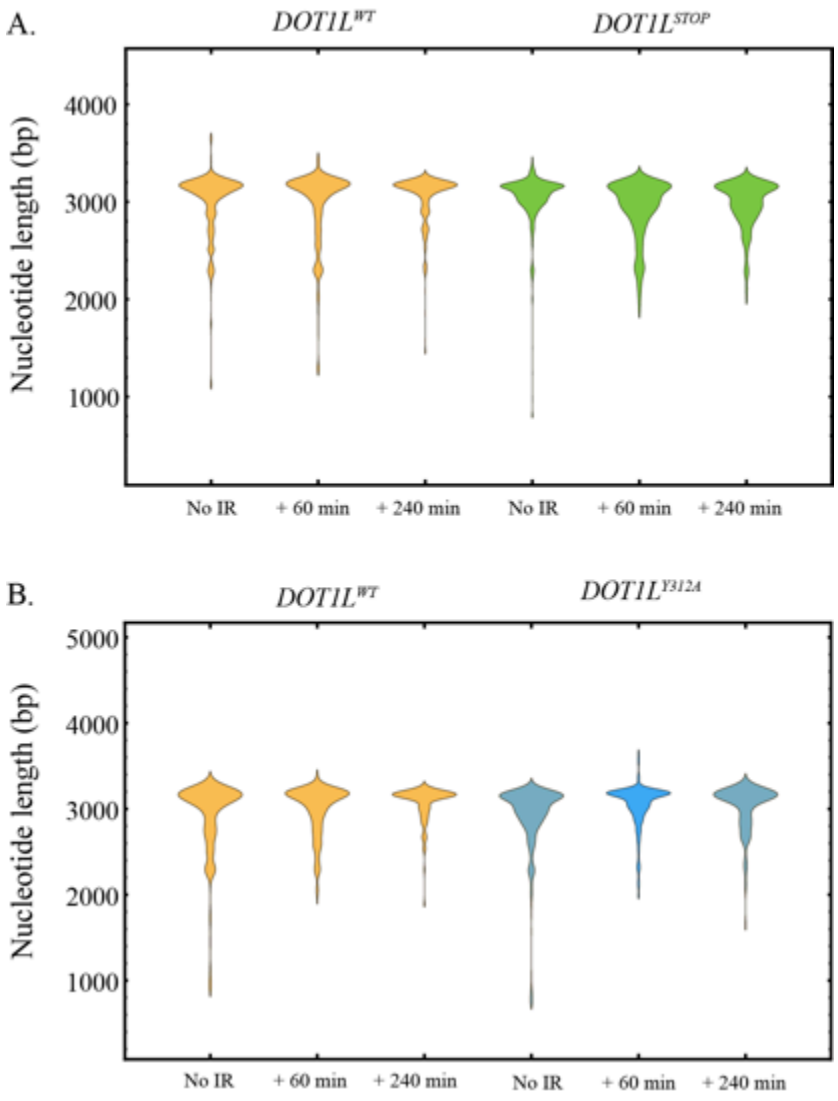


Figure 3.4. ssDNA length distribution for wildtype and *DOT1L* mutants. A. The distribution of ssDNA lengths all the common sequences between wildtype and *DOT1L*^{STOP} mutant cells untreated, or 60 and 240 minutes post radiation treatment. B. The distribution of ssDNA lengths all the common sequences between wildtype and *DOT1L*^{Y312A} mutant cells untreated, or 60 and 240 minutes post radiation treatment.

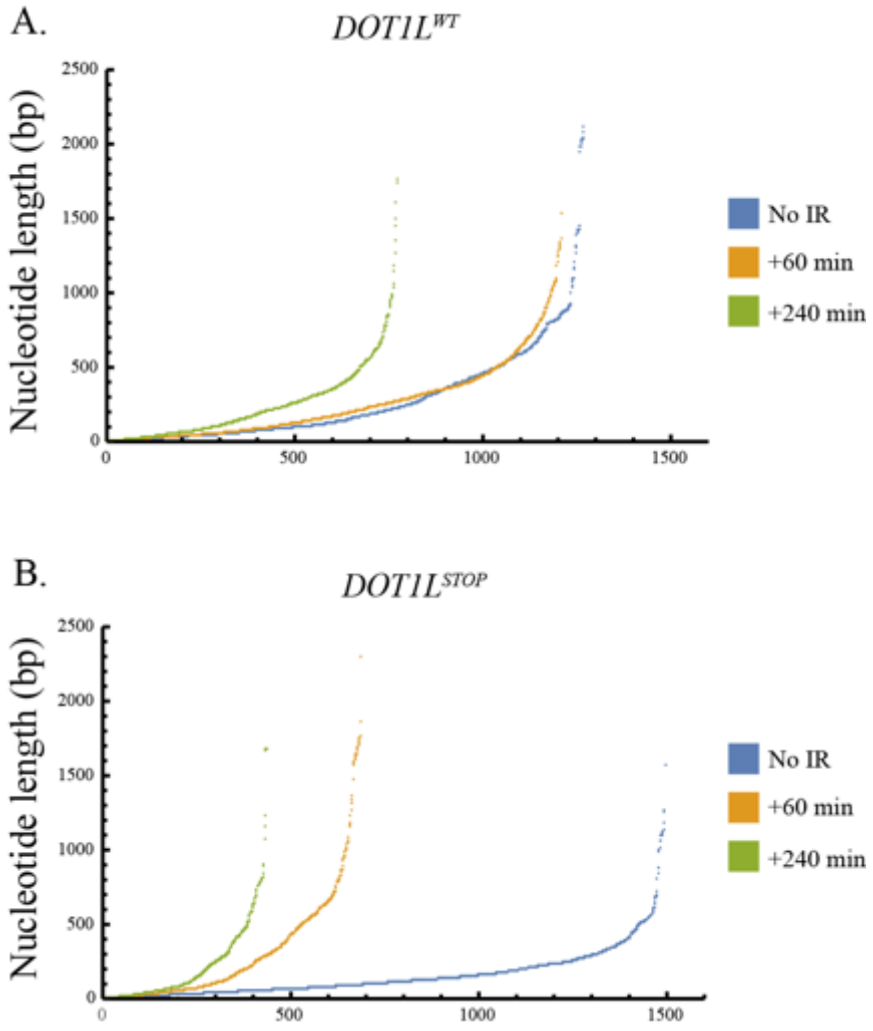


Figure 3.5. *DOT1L*^{STOP} mutant generates longer ssDNA faster than wild-type cells. A. Plot of the difference in length between ssDNA that was identified as longer in the wild-type cells compared to the *DOT1L*^{STOP}. B. Plot of the difference in length between ssDNA that was identified as shorter in the wild-type cells compared to the *DOT1L*^{STOP}.

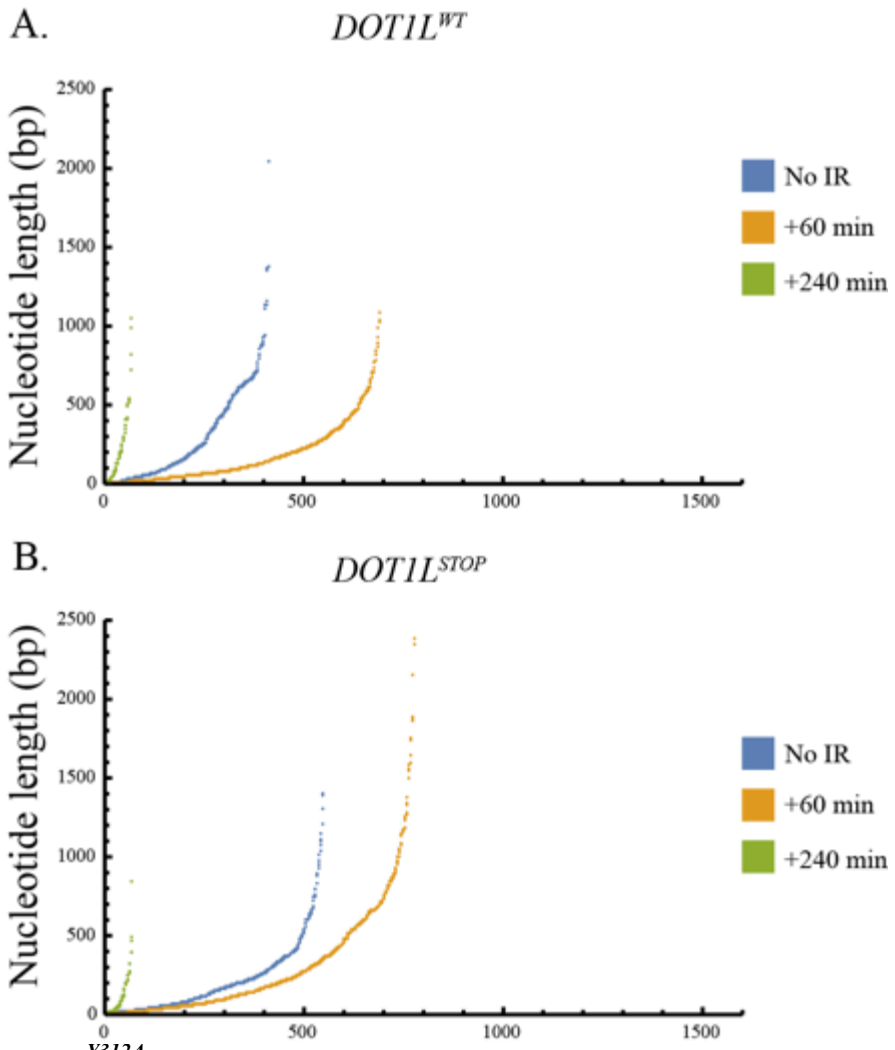


Figure 3.6. *DOT1L*^{Y312A} mutant has longer ssDNA then wild-type cells. A. Plot of the difference in length between ssDNA that was identified as longer in the wild-type cells compared to the *DOT1L*^{Y312A}. B. Plot of the difference in length between ssDNA that was identified as shorter in the wild-type cells compared to the *DOT1L*^{Y312A}.

Chapter IV: Application of a Deep Neural Network for DNA Damage Analysis

Abstract

High-resolution fluorescence microscopy is commonly used to investigate complex dynamic processes such as chromosome organization, transcription factor recruitment, and DNA damage factor localization. High-throughput automated confocal microscopes can generate large amounts of data, yet tools used to analyze the data have remained largely rudimentary. A majority of microscopy image data analysis is still carried out manually. Here, we describe MANA (Machine autonomous nuclei analyzer), an image analysis pipeline used to identify and quantifiy nuclear coalescence events. We developed an analysis pipeline that utilizes machine-learning methods with state-of-the-art 2D and 3D image processing algorithms for the simultaneous staging of the cell cycle and quantification of nuclear localization events. Incorporation of cell cycle data allows for the analysis of asynchronous cell populations and eliminates the need to use exogenous agents for synchronization. We demonstrate the utility of MANA by analyzing the temporal kinetics DNA damage foci localization in human and mouse cells.

Introduction

The various DNA repair pathways present in cells facilitate genome maintenance. Two major DNA repair pathways exist to enable the repair of DNA damage, non-homologous end joining (NHEJ) and homologous recombination (HR). NHEJ is restricted to the G1/G0 phase of the cell cycle, while HR is utilized during S/G2 phase [148]. The evidence for cell cycle dependent DNA repair pathway usage has led to a call for the abandonment of asynchronous cultures in the study of DNA repair pathways [149].

There are technological hurtles to synchronizing cells in culture, and these involve the utilization of drugs that can introduce DNA damage. For example, Aphidicolin, used to synchronize cells in G1/G0 can introduce DSB via replication stress [150]. Etoposide, used to synchronize cells in G2 also introduces DSBs [151]. The use of genotoxic drugs for cell synchronization only confounds the interpretation of data. A more desirable experimental set up is to use asynchronous cultures and separate cells for analysis into cell cycle stages. Escribano-Diaz and collegues utalized the Fucci system in cobination with FACS to study DNA damage reponse in specific phases of the cell cylce [148]. The Fucci system allows for the specific staging of G1 or S/G2 cells using a flouresent proteins fused to CDT1 and Geminin [152]. This system has its limitations with the fluorescent proteins having excitation/emission overlaps with common secondary antibodies used for DNA damage foci analysis. Extensive molecular biology techniques would have to implemented to switch the fluorescent proteins in the Fucci system to ones that would be amenable to co- fluorescent studies. Another confounding issue is the N-terminal domain of Cdt1, used by the Focci system, forms strong interactions with PCNA and participates in the DNA damage respose [153, 154]. Further studies would need to be carried out to determine if overexpression of Cdt1 or Geminin can alter DNA repair pathways.

The use of flow-cytometry with DNA dyes, and flourforescent markers specific for DNA damage factors could be utilized for cell cycle staging, but lack the spacial resolution necessary for the identification of DNA damage foci. Recently a new method for staging cell cycle of cells using a microscopy techiniqe was described [155]. In this technique the authors attempt to overcome many of the limitations mentioined above using a high resolution microscopy in conjuction with DNA dyes to accuarately stage the cell cycle.

Here we present MANA, an image analysis pipeline that uses deep learning neural network and advanced image processing algorithms to analyze cell cycle and nuclear focal events using high-resolution confocal microscopy images. Our method allows for the analysis of asynchronous cell populations with high-resolution nuclear special information (Figure 4.1).

Materials and Methods

Cell Culture

Wild type HEK293T cells were maintained in 10% FBS, DMEM, Pen/Strep in 37 ^{o}C, 5% CO_2.

Microscopy

Microscopy images were collected on a Nikon TE2000 Inverted microscope with a 40X objective, 0.6 NA.

Immunofluorescence

HEK293T cells were plated into 24 well dishes containing L-Poly lysine (75 000 -150 000 MW) treated round cover slips. Cells were exposed to 10 Gy ionizing radiation and fixed at 30 min, 60 min, 90 min, 180 min, 360 minutes and 720 minutes post irradiation with 4% Paraformaldhyde for 30 minutes at room temperature. Cells were washed with 200mM Glycine/1X DPBS for 10 min, permeabilized with 0.2%TritonX100/1XDPBS for 30min, blocked with 1% BSA/1X DPBS for 60 min. 53BP1 and BRCA1 were visualized by staining with primary antibody 53BP1 (1:500) and BRCA1 (1:1000) for 1 hr, followed by secondary antibody AlexFluor 488 (#A-11034, ThermoFisher Scientific) and AlexFluor 594 (#A-11032, ThermoFisher Scientific) at (1:2500) for 2 hours. Cells were washed with 1X DPBS/Hoechst (1µg/mL) 3 times for 5 minutes.

MANA code

```
beaddir =(*directory of bead images*);
beadfiles = Select[FileNames["*", {beaddir}, 1], StringMatchQ[Last[StringSplit[#, "/"]], "*.tif"] &];
(*Imported the data for the TRITC & FITC channel*)
tbimg = Import[#][[1]] & /@ beadfiles;
fbimg = Import[#][[2]] & /@ beadfiles;
dapidir =(*directory for DAPI channel bead images*);
dfiles = Select[FileNames["*", {dapidir}, 1], StringMatchQ[Last[StringSplit[#, "/"]], "*xy1*"] &];
bimg = ColorSeparate[ImageResize[Import[#], Scaled[1/2]]][[3]] & /@ dfiles;

tpsfLocations = ComponentMeasurements[MorphologicalComponents[tbimg[[9]], .1], "Centroid"];
fpsfLocations = ComponentMeasurements[MorphologicalComponents[fbimg[[9]], .1], "Centroid"];
bpsfLocations =
    ComponentMeasurements[SelectComponents[MorphologicalComponents[bimg[[9]], .1], "Area", # < 100 &], "Centroid"];
tzimgs = ImageTrim[#, {tpsfLocations[[All, 2]][[5]]}, 35] & /@ tbimg;
fzimgs = ImageTrim[#, {fpsfLocations[[All, 2]][[5]]}, 35] & /@ fbimg;
bzimgs = ImageTrim[#, {bpsfLocations[[All, 2]][[2]]}, 35] & /@ bimg;
ft = Table[NestList[ImageRotate[#, 5 Degree, Full] &, fzimgs[[i]], 71], {i, 1, Length[fzimgs]}];
ftt = ImageData /@ # & /@ ft;
fout = Image[Mean[#]] & /@ ftt;
tt = Table[NestList[ImageRotate[#, 5 Degree, Full] &, tzimgs[[i]], 71], {i, 1, Length[tzimgs]}];
ttt = ImageData /@ # & /@ tt;
tout = Image[Mean[#]] & /@ ttt;
dt = Table[NestList[ImageRotate[#, 5 Degree, Full] &,
        Threshold[ColorNegate[ImageAdjust[ImageSubtract[BilateralFilter[bzimgs[[i]], 4, 2], bzimgs[[i]]]]], {"Soft", .45}],
        71], {i, 1, Length[bzimgs]}];
dtt = ImageData /@ # & /@ dt;
dout = Image[Mean[#]] & /@ dtt;

dpsf = Image3D[ImageMultiply[ImageCrop[#, 10], 1 / ImageMeasurements[ImageCrop[#, 10], "Total"]] & /@ dout];
fpsf = Image3D[Threshold[ImageCrop[#, 10], {"Soft", .05}] & /@ fout];
tpsf = Image3D[Threshold[ImageCrop[#, 10], {"Soft", .05}] & /@ tout];
isolatenucleimasks[img_, contrast_, brightness_, gamma_, r_, a_, b_, g_, class_] :=
  Block[{d, da, db, distT, marker, w, cell, measures, mz, clustmz, trimcoords,
      imgtrim, mtc, fclustmz, mclustmz, pimg, pimg2, pimg3, pimg4, classify, cpos, cmeas},
    da = ImageAdjust[img, {contrast, brightness, gamma}];
    db = FillingTransform[LocalAdaptiveBinarize[da, r, {a, b, g}]];
    distT = ImageAdjust[DistanceTransform[db, Padding → 0]];
    marker = MaxDetect[distT, 0.02];
    w = WatershedComponents[GradientFilter[db, 3], marker, Method → "Rainfall"];
    cell = SelectComponents[w, "Count", 200 < # < 10000 &];
    measures = ComponentMeasurements[{cell, img}, {"Mask", "Centroid"}];
    pimg = ImageCrop[ImageAdjust[ImageMultiply[Image[#[[2, 1]]], img]]] & /@ measures;
    pimg2 = ImagePad[#, {{Round[50 - ImageDimensions[#][[1]]/2], Round[50 - ImageDimensions[#][[1]]/2]},
          {Round[50 - ImageDimensions[#][[2]]/2], Round[50 - ImageDimensions[#][[2]]/2]}}] & /@ pimg;
    pimg3 = ImageResize[#, Scaled[1/3]] & /@ pimg2;
    pimg4 = ImagePad[#, {{Round[25 - ImageDimensions[#][[1]]/2], Round[25 - ImageDimensions[#][[1]]/2]},
          {Round[25 - ImageDimensions[#][[2]]/2], Round[25 - ImageDimensions[#][[2]]/2]}}] & /@ pimg3;
    classify = class[pimg4];
    cpos = Flatten[Position[classify, 1]];
    cmeas = measures[[#]] & /@ cpos;
    Return[cmeas];]
```

```
dir = SystemDialogInput["Directory"];
files = Select[FileNames["*", {dir}, 1], StringMatchQ[Last[StringSplit[#, "/"]], "*.tif"] &];

ntset = Flatten[isolatenucleimasks[ColorSeparate[ImageResize[Import[#], Scaled[1/2]]][[3]], 0, 0, 1, 10, 0.7, 0, 0.01] & /@ files];

list = {};
Dynamic@list
Table[DynamicModule[{pressed = False}, With[{idx = i}, Button[Show[ntset[[i]], ImageSize → 50], pressed = ! pressed;
        list = If[pressed, Append[list, idx], DeleteCases[list, idx]], Appearance → Dynamic@If[pressed, "Pressed", Automatic]]]],
  {i, 1, Length[ntset]}]

DynamicModule[{idir},
  Button["Isolate Nucli", {idir = SystemDialogInput["Directory"],
    Table[Export[FileNameJoin[{idir, StringJoin["P_", ToString[#], "_", ToString[i], ".tif"]}],
        ImageRotate[ImageResize[ntset[[#]], Scaled[1/3]], i Degree, {50, 50}], "TIFF"], {i, 0, 358, 2}] & /@ list},
    ImageSize → {200, 25},
    Method → "Queued"]]

DynamicModule[{idir},
  Button["Isolate Nucli", {idir = SystemDialogInput["Directory"],
    Table[Export[FileNameJoin[{idir, StringJoin["N_", ToString[#], "_", ToString[i], ".tif"]}],
        ImageRotate[ImageResize[ntset[[#]], Scaled[1/3]], i Degree, {50, 50}], "TIFF"], {i, 0, 358, 2}] & /@
      Keys@Select[Counts[Join[Range[Length[ntset]], list]], # == 1 &]],
    ImageSize → {200, 25},
    Method → "Queued"]]

dir = SystemDialogInput["Directory"];
pfiles = Select[FileNames["*", {dir}, 1], StringMatchQ[Last[StringSplit[#, "/"]], "*.tif"] &];

dir = SystemDialogInput["Directory"];
nfiles = Select[FileNames["*", {dir}, 1], StringMatchQ[Last[StringSplit[#, "/"]], "*.tif"] &];

p = RandomSample[pfiles, 60000];
n = RandomSample[nfiles, 60000];

pData = ParallelTable[Import[p[[i]]] → 1, {i, 1, Length[p]}];
nData = ParallelTable[Import[n[[i]]] → 0, {i, 1, Length[n]}];

data = Flatten[Join[{RandomSample[pData, 60000], RandomSample[nData, 60000]}]];

{traindata, testdata} = TakeDrop[RandomSample[data, 120000], 108000];

lenet = NetChain[{ConvolutionLayer[20, 5], Ramp, PoolingLayer[2, 2],
      ConvolutionLayer[50, 5], Ramp, PoolingLayer[2, 2], FlattenLayer[], 500, Ramp, 2, SoftmaxLayer[]},
    "Output" → NetDecoder[{"Class", Range[0, 1]}], "Input" → NetEncoder[{"Image", {50, 50}, "Grayscale"}]];

lenet = NetTrain[lenet, traindata, ValidationSet → testdata];

covnethek = lenet;
strip3d[xmin_, ymin_, ymax_, zmin_, zmax_] :=
  ImplicitRegion[ymin < y ≤ ymax && xmin − 2 ≤ x ≤ xmin + 2 && zmin < z ≤ zmax, {x, y, z}];

getnucintinsity3d[nuc2dslices_] :=
  Block[{nuc3dmask, nuc3dimg, nx, ny, nz, points, weights, mass, com, pointsCom,
```

```
    rimat, principalaxes, xrot, zrot, rx, rz, rh1, ha1, vi, v1, vi2, va, scan1, ktd2, minktd3, maxktd3,
    minktd2, maxktd2, minktd1, maxktd1, hv, hk, scan3d1, scan3d, v13d, v3d, vi3d, vi23d},
  nuc3dmask = Image3D[LocalAdaptiveBinarize[ImageAdjust[Threshold[#, {"Soft", .017}]], 50, {.85, 0, .01}] & /@
      nuc2dslices];
  nuc3dimg = Image3D[nuc2dslices];
  points = ImageValuePositions[nuc3dmask, 1];
  nx = points[[All, 1]] // Max;
  ny = points[[All, 2]] // Max;
  nz = points[[All, 3]] // Max;
  weights = ImageValue[nuc3dimg, {#[[1]], #[[2]], #[[3]]} & /@ points;
  mass = Total[weights];
  com = Total[points weights]/mass // N;
  pointsCom = (Subtract[#, com] & /@ points) weights;
  (*principle axes*)
  rimat = MomentOfInertia[DiscretizeGraphics[Graphics3D[{Point[pointsCom]}]], com];
  principalaxes = Eigenvectors[rimat];

  xrot = VectorAngle[{1, 0, 0}, principalaxes[[2]]] * 180/Pi;
  zrot = VectorAngle[{0, 0, 1}, principalaxes[[1]]] * 180/Pi;
  rx = RotationTransform[xrot Degree, {{1, 0, 0}, {0, 1, 0}}, com];
  rz = RotationTransform[zrot Degree, {{0, 0, 1}, {1, 1, 0}}, com];

  rh1 = rx[rz[points]];
  ha1 = AssociationThread[rh1, weights];
  ktd2 = Keys[ha1];
  minktd3 = Min[ktd2[[All, 3]]];
  maxktd3 = Max[ktd2[[All, 3]]];
  minktd2 = Min[ktd2[[All, 2]]];
  maxktd2 = Max[ktd2[[All, 2]]];
  minktd1 = Min[ktd2[[All, 1]]];
  maxktd1 = Max[ktd2[[All, 1]]];
  scan3d1 = Table[strip3d[n, minktd2, maxktd2, minktd3, maxktd3], {n, minktd1 - 2, maxktd1 + 2, 0.2}];
  scan3d = Table[Boole[RegionMember[scan3d1[[i]], ktd2]], {i, Length[scan3d1]}];
  v13d = Table[Flatten[Position[scan3d[[i]], 1]], {i, Length[scan3d]}];
  v3d = Table[Values[ha1[[v13d[[i]]]]], {i, Length[v13d]}];
  vi3d = Table[Integrate[Interpolation[Transpose[{Range[Length[v3d[[i]]]], v3d[[i]]}], InterpolationOrder → 0][w],
      {w, 1, Length[v3d[[i]]]}], {i, Length[v3d]}];
  vi23d = Select[vi3d, NumberQ[#] &];
  Return[vi23d];]

nucslicer3d[nucslices_, nucmask_, nuccoords_, psf3d_] :=
 Block[{nuc3d, nuc2d},
  nuc3d = ImageDeconvolve[Image3D[nucslices], psf3d, Method → "RichardsonLucy", MaxIterations → 45];
  nuc2d = ImageMultiply[ImageTrim[Image[nucmask], {nuccoords}, 40], #] & /@ Image3DSlices[nuc3d];
  Return[nuc2d];]

analyzer3d[dslice_, tslice_, fslice_, tdetect_, fdetect_] :=
 Block[{nucint, fticpts, tticpts, coords, coloccount, fitccount, tritccount, r, graphic},
  nucint = getnucintinsity3d[dslice];
  fticpts =
   ComponentMeasurements[{MorphologicalComponents[ImageAdjust[Image3D[fslice]], fdetect, CornerNeighbors → False],
     ImageAdjust[Image3D[fslice]]}, {"Centroid", "Count"}];
  tticpts = ComponentMeasurements[{MorphologicalComponents[ImageAdjust[Image3D[tslice]], tdetect,
```

```
            CornerNeighbors → False], ImageAdjust[Image3D[tslice]]}, {"Centroid", "Count"}];
    coords = Table[Select[Gather[Flatten[{tticpts[[All, 2, 1]], {fticpts[[All, 2, 1]][[i]]}, 1], EuclideanDistance[#1, #2] ≤ 2 &],
        Length[#] ≥ 2 &], {i, 1, Length[fticpts[[All, 2, 1]]]}];
    Return[{nucint, fticpts, tticpts, coords}];]

nucidentify[imgfiles_, dpsf_, tpsf_, fpsf_, r_, a_, b_, g_, class_] :=
    Block[{dmerged, dimg, timg, fimg, dyz, tyz, fyz, d, db, distT, marker, w, cell, measures, pimg, pimg2,
      pimg3, pimg4, classify, cpos, cmeas, dtrim, ttrim, ftrim, dapidecon, tritcdecon, fitcdecon, data, fnames},

      dimg = ColorSeparate[ImageResize[Import[#], Scaled[1/2]]][[3]] & /@ imgfiles;
      timg = ColorSeparate[ImageResize[Import[#], Scaled[1/2]]][[1]] & /@ imgfiles;
      fimg = ColorSeparate[ImageResize[Import[#], Scaled[1/2]]][[2]] & /@ imgfiles;
      dmerged = ImageFocusCombine[dimg];
      db = FillingTransform[LocalAdaptiveBinarize[dmerged, r, {a, b, g}]];
      distT = ImageAdjust[DistanceTransform[db, Padding → 0]];
      marker = MaxDetect[distT, 0.02];
      w = WatershedComponents[GradientFilter[db, 3], marker, Method → "Rainfall"];
      cell = SelectComponents[w, "Count", 200 < # < 10000 &];
      measures = ComponentMeasurements[{cell, dmerged}, {"Mask", "Centroid"}];
      pimg = ImageCrop[ImageAdjust[ImageMultiply[Image[#[[2, 1]]], dmerged]]] & /@ measures;
      pimg2 = ImagePad[#, {{Round[50 − ImageDimensions[#][[1]]/2], Round[50 − ImageDimensions[#][[1]]/2]},
            {Round[50 − ImageDimensions[#][[2]]/2], Round[50 − ImageDimensions[#][[2]]/2]}}] & /@ pimg;
      pimg3 = ImageResize[#, Scaled[1/3]] & /@ pimg2;
      pimg4 = ImagePad[#, {{Round[25 − ImageDimensions[#][[1]]/2], Round[25 − ImageDimensions[#][[1]]/2]},
            {Round[25 − ImageDimensions[#][[2]]/2], Round[25 − ImageDimensions[#][[2]]/2]}}] & /@ pimg3;
      classify = class[pimg4];
      cpos = Flatten[Position[classify, 1]];
      cmeas = measures[[#]] & /@ cpos;
      dtrim = Parallelize[Table[ImageTrim[#, {cmeas[[i, 2, 2]]}, 40] & /@ dimg, {i, 1, Length[cmeas]}]];
      ftrim = Parallelize[Table[ImageTrim[#, {cmeas[[i, 2, 2]]}, 40] & /@ fimg, {i, 1, Length[cmeas]}]];
      ttrim = Parallelize[Table[ImageTrim[#, {cmeas[[i, 2, 2]]}, 40] & /@ timg, {i, 1, Length[cmeas]}]];
      dapidecon = Parallelize[Table[nucslicer3d[dtrim[[i]], cmeas[[i, 2, 1]], cmeas[[i, 2, 2]], dpsf], {i, 1, Length[dtrim]}]];
      fitcdecon = Parallelize[Table[nucslicer3d[ftrim[[i]], cmeas[[i, 2, 1]], cmeas[[i, 2, 2]], fpsf], {i, 1, Length[ftrim]}]];
      tritcdecon = Parallelize[Table[nucslicer3d[ttrim[[i]], cmeas[[i, 2, 1]], cmeas[[i, 2, 2]], tpsf], {i, 1, Length[ttrim]}]];
      fnames = imgfiles;
      DumpSave[FileNameJoin[
          {(*file destination goes here*), StringJoin[StringSplit[fnames, {"/", ".", "_001", "xy", "z"}][[1]][[-5 ;; -3]], ".mx"]}],
        {cmeas, dtrim, ftrim, ttrim, dapidecon, fitcdecon, tritcdecon}];]

counter = 0;
SetSharedVariable[counter];
Monitor[
    Table[counter++;
      nucidentify[(*data*)[[All, j]][[i]], dpsf, tpsf, fpsf, 20, 0.7, 0.1, 0.01, covnethek], {j, 1, Length[(*data*)[[1]]]},
      {i, 1, Length[(*data*)[[All, j]]]}], Row[{ProgressIndicator[counter, {1, Length[Flatten[(*data*), 1]]}], counter}, " "]];

datadir = SystemDialogInput["Directory"];
mxfnames = FileNames["*.mx", {datadir}];

sample = "WT";
times = {"NoIR", "+30min", "+60min", "+90min", "+180min", "+360min", "+720min"};
getdata[mxnames_] :=
  Block[{data},
```

```
    Import[mxnames];
    Clear[cmeas, dtrim, ftrim, ttrim];
    data = ParallelTable[analyzer3d[dapidecon[[i]], tritcdecon[[i]], fitcdecon[[i]], 0.4, 0.2], {i, 1, Length[dapidecon]}];
    Clear[dapidecon, fitcdecon, tritcdecon];
    Return[data];]

search = StringJoin["*", sample, "*", times[[1]], "*"];
mxfiles = Select[mxfnames, StringMatchQ[#, {search}] &];

data = Table[getdata[mxfiles[[i]]], {i, 1, Length[mxfiles]}];

data = Select[Flatten[data[[All, All, 1]], 1], Length[#] > 0 &];
vadjustment[sample_, pw_, thres_] :=
  Block[{transchan, steps, int, bin, height, wtm, crit, wt2m, wt2m2, wtint2, bin2, height2, lh, pr, lh2, pr2, wtint2m,
    wtint3, wtint2m2, wtint4, wtint5, height3, fcoorddata, tcoorddata, ocountdata, fcountdata, tcountdata, limit},
    int = Flatten[Integrate[Interpolation[#][x], {x, 1, Length[#]}] & /@ sample[[All, 1]]];
    wtint4 = Flatten[Max[#] & /@ sample[[All, 1]]];
    fcoorddata = Table[Select[Values[sample[[All, 2]][[i]]], #[[2]] > 10 &], {i, 1, Length[sample[[All, 2]]]}];
    tcoorddata = Table[Select[Values[sample[[All, 3]][[i]]], #[[2]] > 10 &], {i, 1, Length[sample[[All, 3]]]}];
    ocountdata = Table[Length[Select[Gather[Flatten[{tcoorddata[[i]][[All, 1]], fcoorddata[[i]][[All, 1]]}, 1],
          EuclideanDistance[#1, #2] ≤ 2 &], Length[#] ≥ 2 &]], {i, 1, Length[fcoorddata]}];
    fcountdata = Length[#] & /@ fcoorddata;
    tcountdata = Length[#] & /@ tcoorddata;
    {bin, height} = HistogramList[int, {1, 40000, 200}];
    steps = Partition[Range[Length[bin]], {2}, 1];
    height = Table[Select[int, bin[[steps[[i, 1]]]] < # < bin[[steps[[i, 2]]]] &], {i, 1, Length[steps]}];
    height = Transpose[{Length[#] & /@ height, height}];
    transchan = Flatten[Position[bin, 10201]][[1]];
    lh = Round[height[[ ;; Round[N[Length[height]/2]]]]];
    pr = Sort[Flatten[Position[lh[[All, 1]], #] & /@ Range[Max[lh[[All, 1]]] - 0, Max[lh[[All, 1]]]]]];
    wt2m = Round[Mean[{Min[pr], Max[pr]}]];
    wt2m2 = 0;
    crit = {0.};
    limit = {};
    While[
      wt2m2 ≤ transchan + thres,
      wtint2 = Flatten[Integrate[Interpolation[# + Last[crit]][x], {x, 1, Length[#]}] & /@ sample[[All, 1]]];
      wtint3 = Transpose[{wtint2, wtint4, fcountdata, tcountdata, ocountdata}];
      {bin2, height2} = HistogramList[wtint3[[All, 1]], {1, 40000, 200}];
      steps = Partition[Range[Length[bin2]], {2}, 1];
      height2 = Table[Select[wtint3, bin2[[steps[[i, 1]]]] < #[[1]] < bin2[[steps[[i, 2]]]] &], {i, 1, Length[steps]}];
      height2 = Transpose[{Length[#] & /@ height2, height2}];
      lh2 = height2[[ ;; Round[N[Length[height2]/2]]]];
      pr2 = Sort[Flatten[Position[lh2[[All, 1]], #] & /@ Range[Max[lh2[[All, 1]]] - pw, Max[lh2[[All, 1]]]]]];
      wt2m2 = Round[Mean[{Min[pr2], Max[pr2]}]];
      If[Length[Last[Split[limit]]] ≥ 15 && transchan ≤ Last[Split[limit]][[1]] ≤ transchan + thres, Break[]];
      AppendTo[limit, wt2m2];
      AppendTo[crit, If[wt2m2 ≤ transchan - 5 + thres, Last[crit] + 1, Last[crit] + 0.05]];];
    Return[height2];]

g1ctlpeakfinder[sample_, pw_, thres_, sm_] :=
  Block[{int, bins, height, ch, chmin, chmax, shist, lh, muG10, sdG10, sdG10width, lb, ub, chreg, shistreg,
    theta0, lbtheta, ubtheta, fitnorm, theta, muG1, sdG1width, transchan, fp, bins2, height2, shist2, b},
```

```
    int = Select[Flatten[vadjustment[sample, pw, thres]][[All, 2]], 1][[All, 1]], # > 9500 &];
    {bins, height} = HistogramList[int, {1,  40000, 200}];
    ch = Range[Length[height]];
    chmin = 1;
    chmax = Last[ch];
    (*for g1 peak*)
    shist = N[Total[height[[#]] * PDF[NormalDistribution[#, sm], ch] & /@ ch]];
    muG10 = Flatten[Position[shist, Max[shist]]][[1]];
    sdG10 = Flatten[Last[Position[shist, #] & /@ Select[shist, # > 0.6 * Max[shist] &]]][[1]];
    sdG10width = Abs[muG10 - sdG10];
    lb = Max[chmin, First[Select[ch, # > muG10 - 3 * sdG10width &]]];
    ub = First[Select[ch, # > muG10 + sdG10width &]];
    chreg = ch[[lb ;; ub]];
    shistreg = N[shist[[lb ;; ub]]];
    theta0 = Log[#] & /@ {N[muG10], N[sdG10width]};
    lbtheta = Log[#] & /@ {N[lb], N[10^-3]};
    ubtheta = Log[#] & /@ {N[ub], N[(ub - lb)/2]};
    Return[{ch, chmin, shist, height, theta0, lbtheta, ubtheta, chreg, shistreg}];]

histproc[sample_, pw_, thres_] :=
  Block[{int, bins, height, ch, chmin, shist, fp, transchan},
    int = Select[Flatten[vadjustment[sample, pw, thres]][[All, 2]], 1][[All, 1]], # > 9500 &];
    {bins, height} = HistogramList[int, {1,  40000, 200}];
    ch = Range[Length[height]];
    chmin = 1;
    shist = N[Total[height[[#]] * PDF[NormalDistribution[#, .75], ch] & /@ ch]];
    Return[{ch, chmin, shist}];]

g2ctlpeakfinder[ch_, chmin_, shist_, muG1_] :=
  Block[{lb, shistreg, muG20, sdG20, sdG20width, rb,
      ub, chreg, theta0, lbtheta, ubtheta, fitnorm, theta, muG2, sdG2width, shistreg2},
    rb = First[Select[ch, # > 1.75 * muG1 &]];
    shistreg = N[shist][[rb ;;]];
    muG20 = Flatten[Position[N[shist], Max[shistreg]]][[1]];
    sdG20 = Flatten[Last[Position[N[shist], #] & /@ Select[N[shist], # > 0.6 * Max[shistreg] &]]][[1]];
    sdG20width = Abs[sdG20 - muG20];
    lb = Max[chmin, First[Select[ch, # > muG20 - sdG20width &]]];
    ub = First[Select[ch, # > muG20 + 3 * sdG20width &]];
    chreg = ch[[lb ;; ub]];
    shistreg = N[shist][[lb ;; ub]];
    theta0 = N[Log[#] & /@ {muG20, sdG20width};
    lbtheta = Log[#] & /@ {N[lb], N[10^-3]};
    ubtheta = Log[#] & /@ {N[ub], N[(ub - lb)/2]};
    Return[{theta0, lbtheta, ubtheta, chreg, shistreg}];]

sphasectlfinder[ch_, shist_, muG1_, sdG1width_, muG2_, sdG2width_] :=
  Block[{kG1, kG2, crt, G1dens, G2dens, crit1, crit2, Sdens, shistunc,
      theta0, lbtheta, ubtheta, fitsumnorm, theta, MuG1, SDG1width, MuG2, SDG2width},
    kG1 = -3.;
    kG2 = -3.;
    crit1 = {0.};
    G1dens = N[PDF[NormalDistribution[muG1, sdG1width], ch]];
    G2dens = N[PDF[NormalDistribution[muG2, sdG2width], ch]];
```

```
    While[
      Erf[kG1] ≤ Last[crit1],
      Sdens = 0.5 * (Erf[(((ch - muG1)/sdG1width) - kG1)] - Erf[(((ch - muG2)/sdG2width) + kG2)]);
      kG1 = kG1 + 0.05;
      AppendTo[crit1, N[Sdens[[muG1]]/(2 * G1dens[[muG1]])]];];
    crit2 = {0.};
    While[
      Erf[kG2] ≤ Last[crit2],
      Sdens = 0.5 * (Erf[(((ch - muG1)/sdG1width) - kG1)] - Erf[(((ch - muG2)/sdG2width) + kG2)]);
      kG2 = kG2 + 0.05;
      AppendTo[crit2, N[Sdens[[muG2]]/(2 * G2dens[[muG2]])]];];
    shistunc = shist - shist * Sdens;
    theta0 = N[Log[#]] & /@ {N[muG1], N[sdG1width], N[muG2], N[Max[{sdG1width, sdG2width}]]};
    lbtheta = N[Log[#]] & /@ Flatten[{N[10^-3]} & /@ Range[4]];
    ubtheta = N[Log[#]] & /@ {N[Max[ch]], N[Max[ch]], N[Max[ch]], N[sdG1width]};
    Return[{shistunc, theta0, lbtheta, ubtheta, Sdens}];]
wtnodata2 = Select[Flatten[wtNoIR, 1], Length[#[[1]]] > 0 &];
wt30data2 = Select[Flatten[wt30, 1], Length[#[[1]]] > 0 &];
wt60data2 = Select[Flatten[wt60, 1], Length[#[[1]]] > 0 &];
wt720data2 = Select[Flatten[wt720, 1], Length[#[[1]]] > 0 &];

wtdata = {{wtnodata2, 25}, {wt30data2, 30}, {wt60data2, 30}, {wt720data2, 50}};
wtvadj = vadjustment[#[[1]], 0, #[[2]]] & /@ wtdata;
getfocipct[data_, g1l_, g1u_, g2l_, g2u_, foci_, cutoff_] :=
  Block[{g1f, sf, g2f, g1p, sp, g2p},
    g1f = Length[Select[Flatten[data[[g1l ;; g1u]]][[All, 2]], 1][[All, foci]], # > cutoff &]];
    sf = Length[Select[Flatten[data[[g1u ;; g2l]]][[All, 2]], 1][[All, foci]], # > cutoff &]];
    g2f = Length[Select[Flatten[data[[g2l ;; g2u]]][[All, 2]], 1][[All, foci]], # > cutoff &]];
    g1p = (g1f/(g1f + sf + g2f)) * 100 // N;
    sp = (sf/(g1f + sf + g2f)) * 100 // N;
    g2p = (g2f/(g1f + sf + g2f)) * 100 // N;
    Return[{g1p, sp, g2p}];]
ecscatter[data_, intpad_, g1l_, g1u_, g2l_, g2u_, foci_, cutoff_] :=
  Show[ListPlot[{Transpose[{Flatten[data[[All, 2]], 1][[All, 1]],
            intpad + Flatten[data[[All, 2]], 1][[All, 2]], Flatten[data[[All, 2]], 1][[All, foci]]}][[All, 1 ;; 2]],
      Select[Transpose[{Flatten[data[[All, 2]], 1][[All, 1]], intpad + Flatten[data[[All, 2]], 1][[All, 2]],
          Flatten[data[[All, 2]], 1][[All, foci]]}], #[[3]] ≥ cutoff &][[All, 1 ;; 2]]},
    Epilog → {Black, Dashed, Line[{{Min[Flatten[data[[g1l ;; g1u]]][[All, 2]], 1][[All, 1]]], 0},
        {Min[Flatten[data[[g1l ;; g1u]]][[All, 2]], 1][[All, 1]]], 120}}],
      Line[{{Max[Flatten[data[[g1l ;; g1u]]][[All, 2]], 1][[All, 1]]], 0},
        {Max[Flatten[data[[g1l ;; g1u]]][[All, 2]], 1][[All, 1]]], 120}}],
      Line[{{Min[Flatten[data[[g2l ;; g2u]]][[All, 2]], 1][[All, 1]]], 0}, {Min[Flatten[data[[g2l ;; g2u]]][[All, 2]], 1][[All, 1]]],
        120}}], Line[{{Max[Flatten[data[[g2l ;; g2u]]][[All, 2]], 1][[All, 1]]], 0},
        {Max[Flatten[data[[g2l ;; g2u]]][[All, 2]], 1][[All, 1]]], 120}}]},
    PlotStyle → {Automatic, Red}, PlotRange → {{0, 45 000}, {0, 120}}],
    Graphics[{Text[Style[getfocipct[data, g1l, g1u, g2l, g2u, foci, cutoff], Medium], {43 000, 10}, {1, 0}]}]]
{ch, chmin, shist, height, theta01, lbtheta1, ubtheta1, chreg1, shistreg1} = g1ctlpeakfinder[wtnodata2, 0, 25, .75];

fitnorm[a_, b_] :=
  Total[PDF[NormalDistribution[a, b], chreg1]/PDF[NormalDistribution[a, b], a] - (shistreg1/Max[shistreg1])]

theta1 = FindMinimum[fitnorm[Exp[a], Exp[b]],
    {{a, theta01[[1]], lbtheta1[[1]], ubtheta1[[1]]}, {b, theta01[[2]], theta01[[2]], ubtheta1[[2]]}}];
muG1 = Round[Exp[theta1[[2, 1, 2]]]];
```

```
sdG1width = Round[Exp[theta1[[2, 2, 2]]]];

{theta02, lbtheta2, ubtheta2, chreg2, shistreg2} = g2ctlpeakfinder[ch, chmin, shist, muG1];

fitnorm[a_, b_] :=
  Total[PDF[NormalDistribution[a, b], chreg2] / PDF[NormalDistribution[a, b], a] - (shistreg2 / Max[shistreg2])]

theta2 = FindMinimum[fitnorm[Exp[a], Exp[b]],
      {{a, theta02[[1]], lbtheta2[[1]], ubtheta2[[1]]}, {b, theta02[[2]], theta02[[2]], ubtheta2[[2]]}}];

muG2 = Round[Exp[theta2[[2, 1, 2]]]];
If[muG2 > 2 * muG1, muG2 = muG1 * 2];
sdG2width = Round[Exp[theta2[[2, 2, 2]]]];

{sdG1width, sdG2width} = {10, 15};

{shistunc, theta03, lbtheta3, ubtheta3, Sdens} = sphasectlfinder[ch, shist, muG1, sdG1width, muG2, sdG2width];

fitsumnorm[a_, b_, c_, d_] :=
  Total[((Indexed[shistunc, Round[a]] / PDF[NormalDistribution[a, b], a]) * PDF[NormalDistribution[a, b], ch]) +
      (PDF[NormalDistribution[c, d], ch] * (Indexed[shistunc, Round[c]] / PDF[NormalDistribution[c, d], c])) - shistunc]

theta3 = FindMinimum[fitsumnorm[Exp[a], Exp[b], Exp[c], Exp[d]],
      {{a, theta03[[1]], lbtheta3[[1]], ubtheta3[[1]]}, {b, theta03[[2]], theta03[[2]], ubtheta3[[2]]},
        {c, theta03[[3]], lbtheta3[[3]], ubtheta3[[3]]}, {d, theta03[[4]], theta03[[4]], ubtheta3[[4]]}}];

MuG1 = Round[Exp[theta3[[2, 1, 2]]]
SDG1width = Round[Exp[theta3[[2, 2, 2]]]]
MuG2 = Round[Exp[theta3[[2, 3, 2]]]]
If[MuG2 > 2 * MuG1, MuG2 = MuG1 * 2];
SDG2width = Round[Exp[theta3[[2, 4, 2]]]]

r = {Min[#], Max[#]} & /@ Split[Flatten[Position[Round[shist - shistunc], Min[Round[shist - shistunc]]]], #2 - #1 == 1 &];
smid = Round[N[Mean[Range[First[r][[2]], Last[r][[1]]]]]];
s = shist - shistunc;
smaxp = Flatten[Position[(shist - shistunc), Max[(shist - shistunc)]]][[1]];

rg = {Min[#], Max[#]} & /@
      Split[Flatten[Position[Boole[# > 0 & /@ Round[shistunc]], Max[Boole[# > 0 & /@ Round[shistunc]]]]], #2 + #1 == 2 &];
gwd = Select[{Min[#], Max[#]} & /@ Split[Select[rg[[All, 1]], # > 2 &], #2 - #1 ≤ 10 &], #[[2]] - #[[1]] ≥ 2 &]
g1wd = First[gwd];
If[Length[gwd] == 1, g2 = 0, g2wd = gwd[[2]]];
rg2 = {Min[#], Max[#]} & /@
      Split[Flatten[Position[Boole[# > 0 & /@ Round[shist - shistunc]], Max[Boole[# > 0 & /@ Round[shist - shistunc]]]]],
        #2 + #1 == 2 &];
swd = First[Select[{Min[#], Max[#]} & /@ Split[rg2[[All, 1]], #2 - #1 == 1 &], #[[2]] - #[[1]] ≥ 5 &]];

g1 = Total[shistunc[[g1wd[[1]] ;; g1wd[[2]]]]];
s2 = Total[(shist - shistunc)[[swd[[1]] ;; swd[[2]]]]];
If[Length[gwd] == 1, g2 = 0, g2 = Total[shistunc[[g2wd[[1]] ;; g2wd[[2]]]]]];
prct = N[{{g1 / (g1 + g2 + s2), s2 / (g1 + g2 + s2), g2 / (g1 + g2 + s2)}} * 100];

Show[ListLinePlot[{shist * 0}, PlotRange → {{1, 300}, {0, 20}}, AxesStyle → {Black, Thickness[.005]}],
  BarChart[{shist}, ChartStyle → Black, PlotRange → {{0, 300}, {0, 25}}],
```

```
  ListLinePlot[{shistunc[[MuG1]]/PDF[NormalDistribution[MuG1, SDG1width], MuG1]*
        PDF[NormalDistribution[MuG1, SDG1width], ch], shistunc[[MuG2]]/
          PDF[NormalDistribution[MuG2, SDG2width], MuG2]*PDF[NormalDistribution[MuG2, SDG2width], ch],
      s[[smaxp]]/PDF[NormalDistribution[smaxp, SDG1width], smaxp]*PDF[NormalDistribution[smaxp, SDG1width], ch]},
    PlotRange → {{0, 300}, {0, 25}}, PlotStyle → Thickness[.01], PlotLegends → {"G1", "G2", "S"}],
  Graphics[{Text[Style[prct[[1, 1]], Medium], {250, 6}, {1, 0}], Text[Style[prct[[1, 2]], Medium], {250, 4}, {1, 0}],
      Text[Style[prct[[1, 3]], Medium], {250, 2}, {1, 0}]}]]
{kch, kchmin, kshist} = histproc[wt60data2, 0, 30];

{kshistunc, ktheta03, klbtheta3, kubtheta3, kSdens} = sphasectlfinder[kch, kshist, MuG1, SDG1width, MuG2, SDG2width];

theta3 = FindMinimum[fitsumnorm[Exp[a], Exp[b], Exp[c], Exp[d]],
      {{a, ktheta03[[1]], ktheta03[[1]], ktheta03[[1]]}, {b, ktheta03[[2]], ktheta03[[2]], kubtheta3[[2]]},
        {c, ktheta03[[3]], ktheta03[[3]], ktheta03[[3]]}, {d, ktheta03[[4]], ktheta03[[4]], kubtheta3[[4]]}}];

kMuG1 = Round[Exp[theta3[[2, 1, 2]]]]
kSDG1width = Round[Exp[theta3[[2, 2, 2]]]]
kMuG2 = Round[Exp[theta3[[2, 3, 2]]]]
If[kMuG2 > 2*kMuG1, kMuG2 = kMuG1*2]
kSDG2width = Round[Exp[theta3[[2, 4, 2]]]]

kr =
    {Min[#], Max[#]} & /@ Split[Flatten[Position[Round[kshist - kshistunc], Min[Round[kshist - kshistunc]]]], #2 - #1 == 1 &];

ksmid = Round[N[Mean[Range[First[kr][[2]], Last[kr][[1]]]]]];

ks = kshist - kshistunc;
ksmaxp = Flatten[Position[(kshist - kshistunc), Max[(kshist - kshistunc)]]][[1]];

krg = {Min[#], Max[#]} & /@
      Split[Flatten[Position[Boole[# > 0 & /@ Round[kshistunc]], Max[Boole[# > 0 & /@ Round[kshistunc]]]]], #2 + #1 == 2 &];
kgwd = Select[{Min[#], Max[#]} & /@ Split[Select[krg[[All, 1]], # > 2 &], #2 - #1 ≤ 10 &], #[[2]] - #[[1]] ≥ 2 &]
kg1wd = First[kgwd];
If[Length[kgwd] == 1, kg2 = 0, kg2wd = Last[kgwd]];
krg2 = {Min[#], Max[#]} & /@
      Split[Flatten[Position[Boole[# > 0 & /@ Round[kshist - kshistunc]], Max[Boole[# > 0 & /@ Round[kshist - kshistunc]]]]],
        #2 + #1 == 2 &];
kswd = First[Select[{Min[#], Max[#]} & /@ Split[krg2[[All, 1]], #2 - #1 == 1 &], #[[2]] - #[[1]] ≥ 10 &]];

kg1 = Total[kshistunc[[kg1wd[[1]] ;; kg1wd[[2]]]]];
ks2 = Total[(kshist - kshistunc)[[kswd[[1]] ;; kswd[[2]]]]];
If[Length[kgwd] == 1, kg2 = 0, kg2 = Total[kshistunc[[kg2wd[[1]] ;; kg2wd[[2]]]]]];
kprct = N[{{kg1/(kg1 + kg2 + ks2), ks2/(kg1 + kg2 + ks2), kg2/(kg1 + kg2 + ks2)}}*100];

Show[ListLinePlot[{kshist*0}, PlotRange → {{1, 300}, {0, 30}}, AxesStyle → {Black, Thickness[.005]}],
  BarChart[{kshist}, ChartStyle → Black, PlotRange → {Automatic, {0, 30}}, AxesStyle → {Black, Thickness[.005]}],
  ListLinePlot[{kshistunc[[MuG1]]/PDF[NormalDistribution[MuG1, kSDG1width], MuG1]*
        PDF[NormalDistribution[MuG1, kSDG1width], ch],
      kshistunc[[MuG2]]/PDF[NormalDistribution[MuG2, kSDG2width], MuG2]*
        PDF[NormalDistribution[MuG2, kSDG2width], ch], ks[[ksmaxp]]/
          PDF[NormalDistribution[ksmaxp, kSDG1width], ksmaxp]*PDF[NormalDistribution[ksmaxp, kSDG1width], ch]},
    PlotRange → {Automatic, {0, 30}}, PlotStyle → Thickness[.01], PlotLegends → {"G1", "G2", "S"}],
  Graphics[{Text[Style[kprct[[1, 1]], Medium], {250, 6}, {1, 0}], Text[Style[kprct[[1, 2]], Medium], {250, 4}, {1, 0}],
      Text[Style[kprct[[1, 3]], Medium], {250, 2}, {1, 0}]}]]
```

Results

Our implementation of MANA relies on the Mathematica programing language. Mathematica was chosen because of its robust and integrated machine learning and image analysis algorithms that facilitate a homogenous software development environment without the need for prior knowledge of multiple programing languages. It also allows for the generation of graphical user interface that facilitates the dissemination of MANA to individuals with no computer programing language experience.

MANA relies heavily on its ability to determine what constitutes a nucleus in an image. Image segmentation algorithms are used to isolate objects within the images, and a deep learning neural network is used to classify the objects as either nuclei or non-nuclei. Image segmentation relies on an algorithm that determines the local binary threshold values between the foreground and background in an image.

We developed an image segmentation pipeline that relies on local binary threshold rather than a global threshold algorithm to identify individual cells (Figure 4.2). Local thresholding allows for resolving of edges in images where objects are closely spaced, and is superior to global thresholding where a single threshold is picked for the entire image. The application of a global threshold results in erosion of nuclei due to uneven illumination across the image using a local adaptive binarization technique that allows for the capture of nuclei edges more robustly (Figure 4.3, Figure 4.4d).

We compared a local thresholding algorithm to a global thresholding algorithm in our segmentation pipeline. Both local and global algorithms can identify a similar number of objects and locations (Figure 4.4a,b,c). However, global thresholding results in the erosion of borders of

nuclei (Figure 4.4d). In order to maximize the area of nuclei recovered we utilized a local binary thresholding algorithm for the segmentation pipeline.

Having determined the optimal settings for image segmentation, we built a training database of objects curated from microscopy images in order to train a deep neural network to classify objects into “nuclei” or “non-nuclei”. To expand the image training data set, human identified nuclei and non-nuclei were subjected to a two degree shift in orientation to generate a new training set of 189,360 images (Figure 4.5). The training set was randomly sampled to generate a sub list of 120,000 images that was used to train a deep neural network (Figure 4.5). The network achieved 99.55% accuracy with less than 120 minutes of training (Figure 4.5).

MANA can process single plane images as well as images from a z-stack. A z-stack represents multiple optical slices through an object that allow for more precise observation of nuclear localization events. However z-stacking introduces optical artifacts caused by the recording of multiple diffraction patterns from multiple point sources of light. The point source emits light over a wide area, as a result a point light source will balloon in diameter as the focal plane moves farther from the source in either the positive or negative z direction. To compensate for this, images were deconvoluted. Deconvolution is a technique that removes artifacts by applying a point spread function to the image. Typically, this is done using a theoretical point spread function, however, each microscope has its own unique point spread function as a result of minute differences in optical paths. Therefore, we designed MANA to incorporate an experimental point spread function.

We experimentally determined the point spread function for our microscope system using two methods: first was a bead based method, and second was a search of our images for a point source of light. For our first method we used 3.0-3.5 μm rainbow beads designed for flow

cytometry laser calibration. The diameter of the beads was too wide for calculating the point spread function, however the sizes of the beads is not uniform and many sub-micron diameter beads exist, and we used these for a point source for FITC and TRITC channels (Figure 4.6a). The fluorescence in the DAPI channel was extremely weak and the beads could not be used for this channel. Therefore we implemented our second approach and scanned our nuclei images for sub-micron point sources of light in the DAPI channel (Figure 4.6a). We isolated the point sources of light for the TRITC, FITC, and DAPI channels and subjected them to circular symmetry processing to generate a diffraction pattern for each channel (Figure 4.6b). Image representations of the point spread function for each channel were generated by taking the maximum value in the y direction of each z stack image after circular symmetry processing (Figure 4.6c).

We applied the Richardson-Lucy deconvolution algorithm using the experimental point spread function for each channel in the z-stack [156, 157]. Deconvolution is computationally intensive, therefore deconvolution is only applied to isolated nuclei rather than the whole image. Figure 4.7, 4.8, 4.9 demonstrates the before and after images of deconvolution to a single nuclei for the DAPI, FITC and TRITC channels. The deconvoluted z-stack images are combined to generate a 3D representation of the nucleus (Figure 4.10a). Using an image segmentation technique that separates foreground pixels from background pixels identifies bright spots in in the FITC and TRITC channels (Figure 4.10b). The coordinates for nuclei foci are extracted and plotted onto graph for better visualization, and quantification (Figure 4.11).

Next, we performed a cell cycle staging into MANA. The fluorescence intensity of nuclear dyes such as Hoechst are directly linked to the amount of DNA present in the nucleus. Therefore it is possible to segregate nuclei into different groups, with 2N to 4N amount of DNA.

We designed a robust method to quantify DNA dye intensity using optical sections from a z-stack.

First, the 3D nuclei image is converted to a plot representation of the image and then converted to a mass distribution representation of the image (Figure 4.12a,b). The mass distribution representation allows for the calculation of the center of mass, moment of inertia, and principle axes (Figure 4.12c). The degree difference between the principle axes and the xyz plane of zero rotation is used to calculate the rotational angle needed to rotate the nucleus such that the principle axes is perpendicular to the x axis (Figure 4.12d,e). We then collect pixel intensity values by applying a scanning window across the nucleus (Figure 4.13a). The area under the curve is calculated for each sliding window to generate an intensity profile (Figure 4.13b).

From the intensity profile of each nuclei we calculate the area under the curve and plot it verses the maximum intensity value to generate a scatter plot that can separate nuclei into distinct clusters (Figure 4.14a). To determine which intensities correspond to phases of the cell cycle we implemented the Watson method for cell cycle peak calling [158, 159].

We applied MANA to quantify DNA damage repair foci formation and cell cycle staging HEK293T cells that were treated with 10Gy of ionizing radiation. We analyzed the cell cycle of cells 30 minutes, 60 minutes, and 720 minutes post irradiation exposure. MANA provides similar cell cycle staging as that has been reported for H3K293T cells exposed to ionizing radiation (Figure 4.14) [160]. We did observe lower than expected G2/M percentages. This is a technical issue with cell handling for HEK293T cells and not a limitation of MANA. G2/M phase HEK293T cells are easily removed from plates with little agitation. Therefor, it is necessary to

handle cultures carefully during sample preparation. Our experiments with embryonic mouse fibroblast yielded better results with respect to G2/M population representation (data not shown).

The most powerful aspect of MANA is its capability to incorporate high-resolution nuclear foci information for each cell in each phase of the cell cycle (Figure 4.15,16,17). This allows for precise measurements of spetial and temporal nuclear events in high resolution. The benefit for incorporating high-resolution microscopy into cell cycle data is it allows for the proper identification of nuclear foci. Widely used flow cytometers are incapable of distinguishing diffuse from punctate staining. Diffuse staining patterns of dyes and markers within nuclei is a common problem and can lead to misinterpretation of biological significance of staining. MANA can process high-resolution microscopy images in a high-throughput autonomous fashion allowing for the design and implementation of experiments that would otherwise be time prohibitive and impossible to control for user bias analysis.

Discussion

We applied MANA to measure cell cycle kinetics and DNA damage foci. The use of MANA is not limited to only DNA damage foci. Any cellular event that leads to the coalescence of proteins, for which antibodies exist, can be analyzed by MANA. This technology allows for the high throughput analysis of high resolution microscopy images in a nearly machine autonomous fashion. As deep learning methodology improves and computers become faster we predict that the wide spread adoption deep learning neural networks in a wide variety of experimental data analysis will become commonplace. Areas of research that rely on microcopy for experimentation stand to benefit immediately.

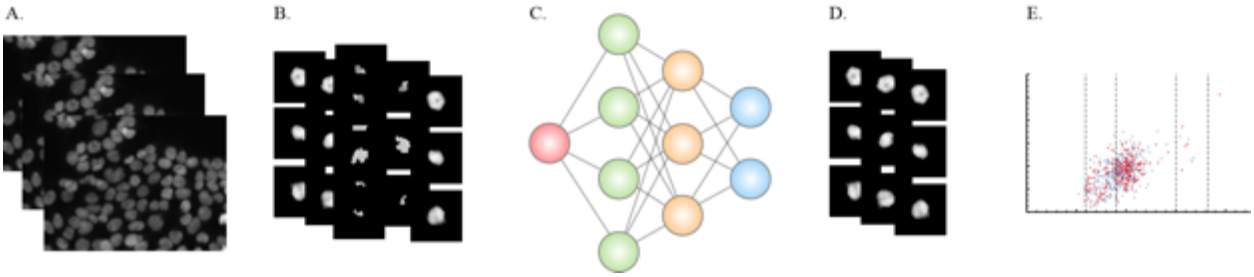


Figure 4.1: MANA pipeline overview. A. High resolution microscopy images are (B) image segmented to isolate objects (C) and these objects are analyzes by a deep learning neural network to (D) identify nuclei and subsequent steps quantify (E) foci amount and cycle stage.

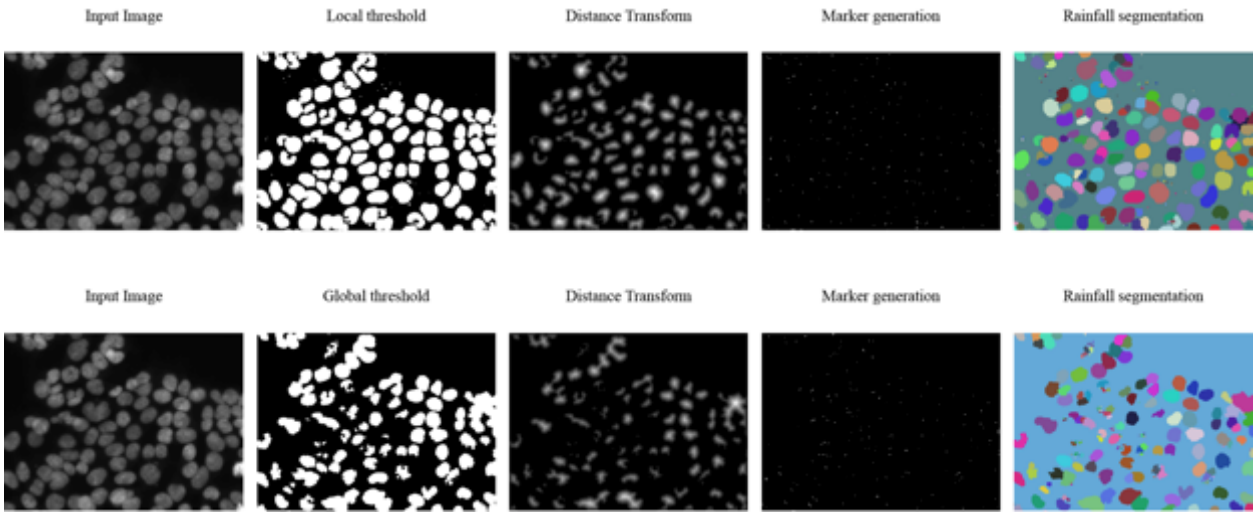


Figure 4.2: Image segmentation using local binarization. The images displayed in the top row are the results of segmentation of nuclei using a thresholding algorithm that acts locally. The images on the bottom row are the results of segmentation of nuclei using a global thresholding algorithm. The sequential output of the segmentation pipeline is left to right.

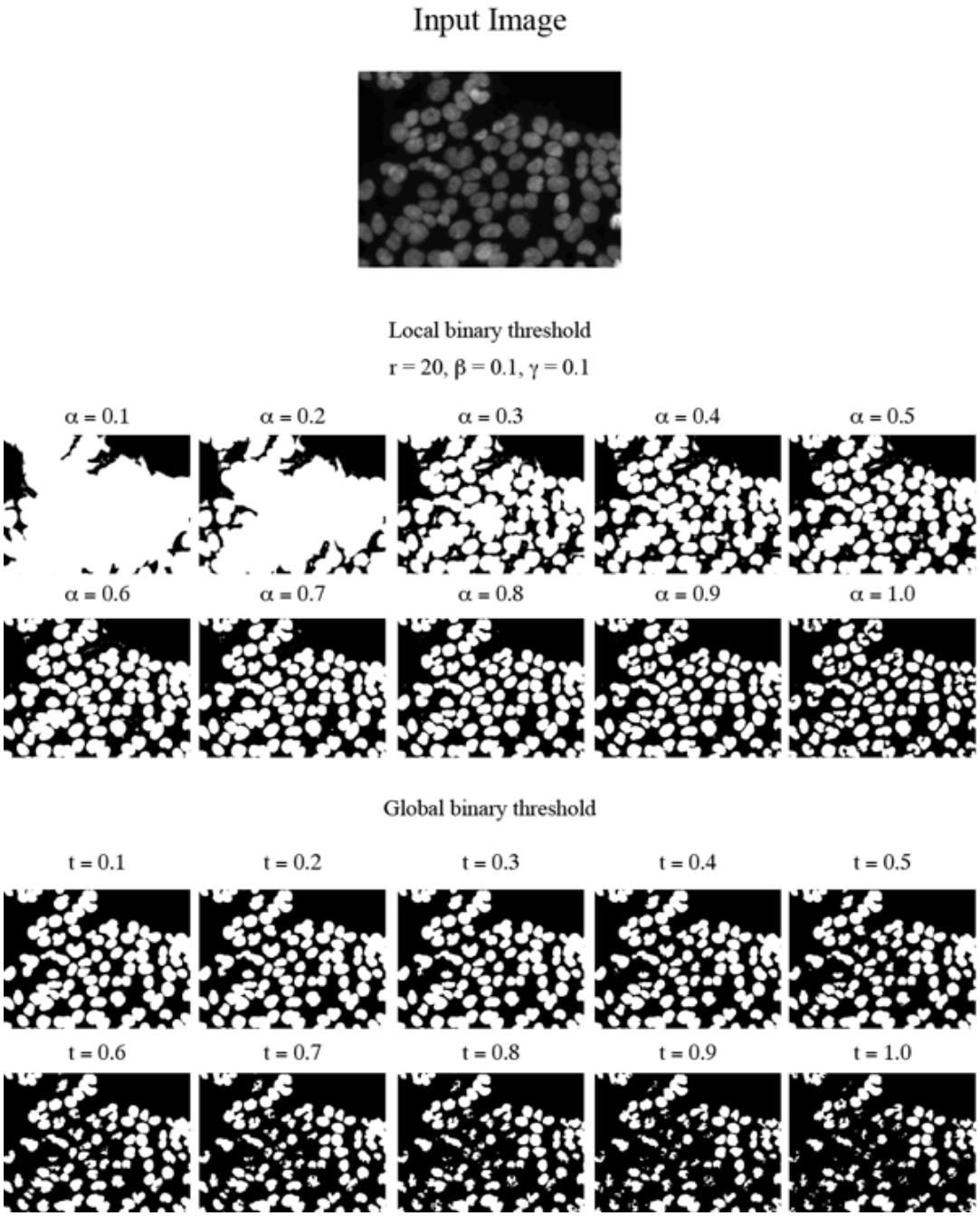


Figure 4.3: Comparison of local and global binary threshold algorithms. (Top, from left to right) Local binary threshold with various settings that demonstrate the ability to resolve closely spaced object without erosion. (Bottom, from left to right) Global binary threshold with various settings demonstrate that the algorithm results in a high level of erosion of object.

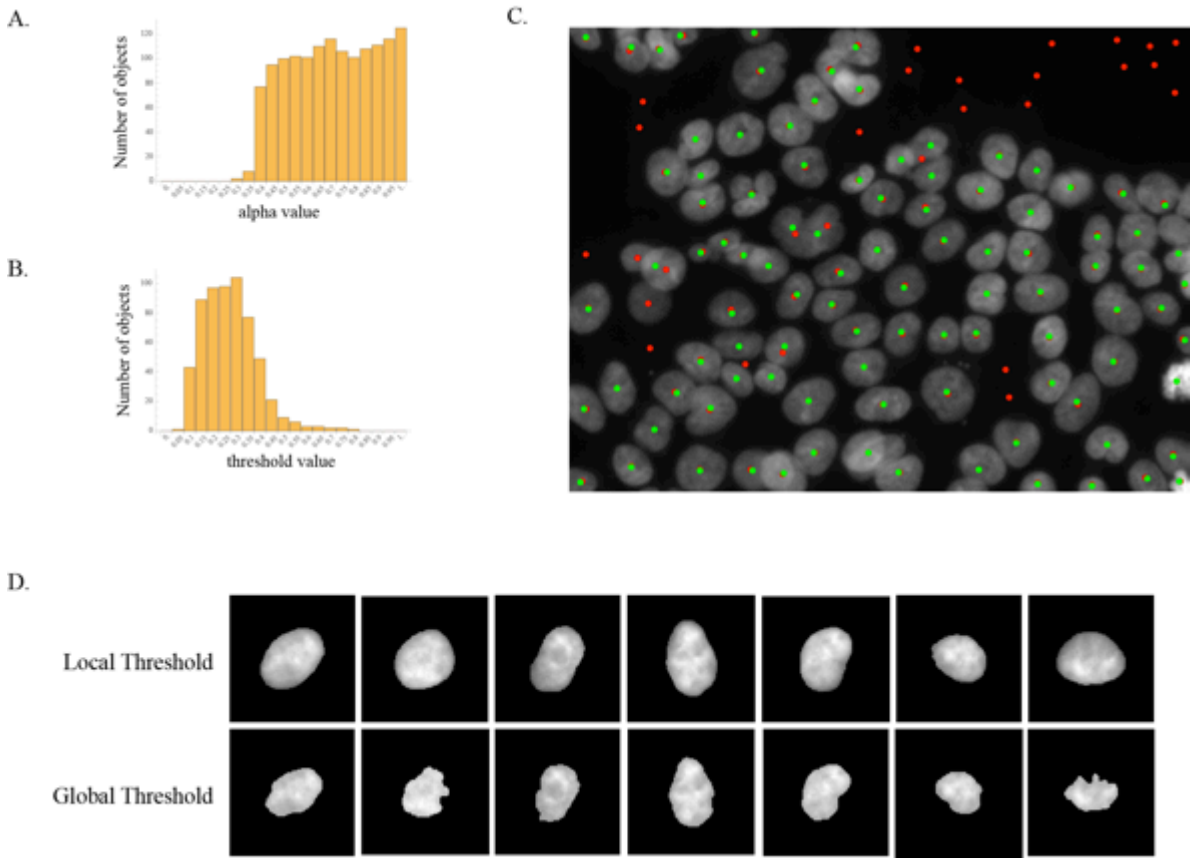


Figure 4.4: Object identification using local and global threshold algorithms. A. Number of objects identified by local thresholding at various settings. B. Number of object identified by global thresholding at various settings. C. (Red) local threshold object, (Green) global threshold objects. D. Visualization of the erosion of nuclei boundaries using global thresholding.

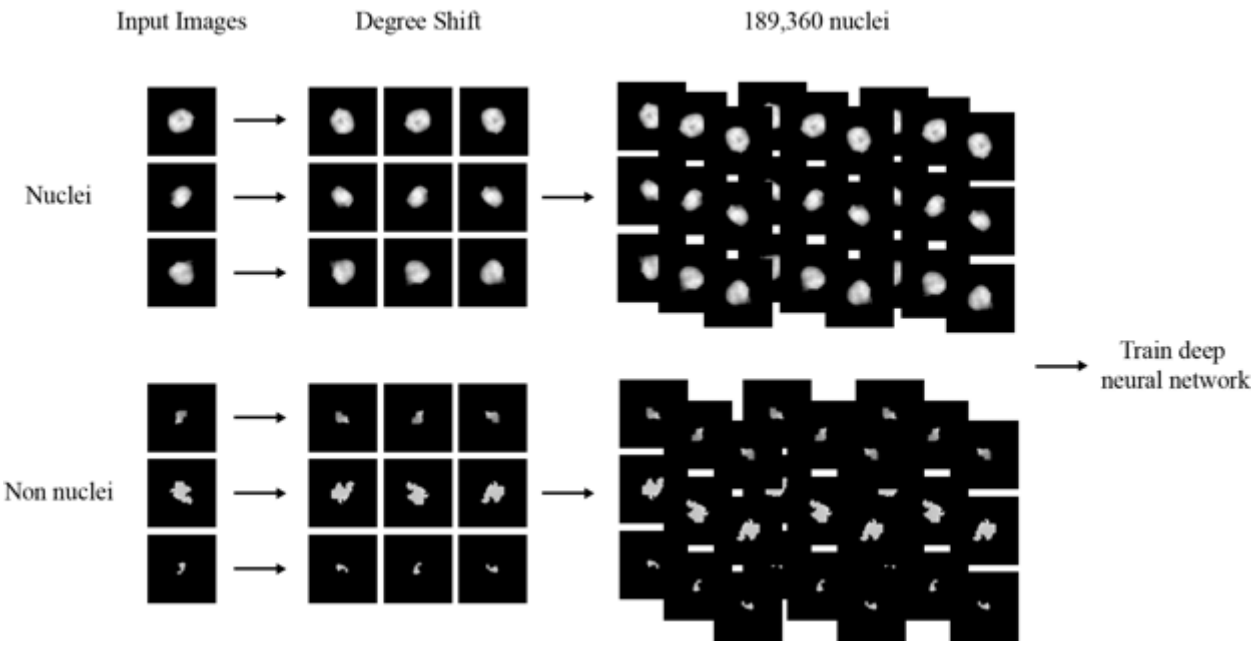


Figure 4.5: Deep neural network training for nuclei identification. Segmentation analysis allows for the isolation of objects that could be classified as nuclei or non nuclei. A human curated list of nuclei and non nuclei was expanded using a degree shift procedure in which each object was rotated 2 degrees. The final data set contains over 180,000 nuclei and non nuclei object that is used to train a deep neural network.

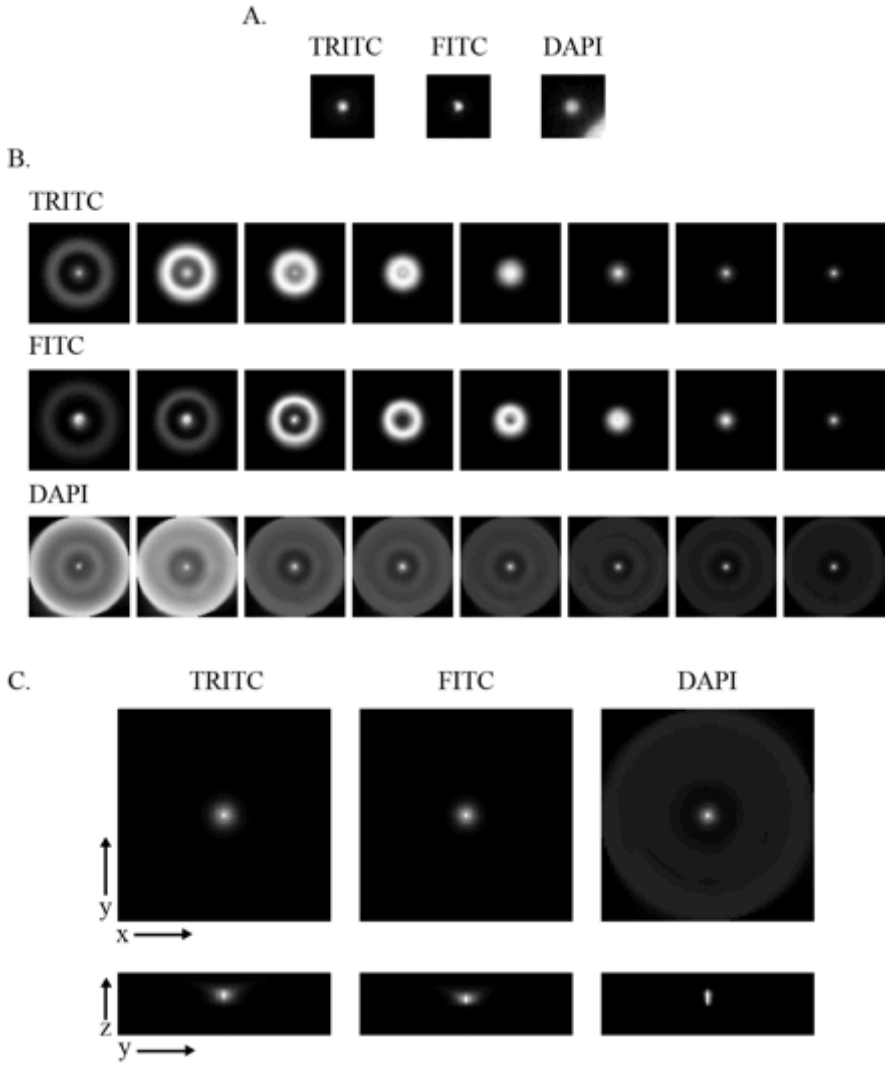


Figure 4.6: Experimental point spread function calculation. A. Point sources of light were isolated from images using TRITC, FITC, and DAPI channels. B. Isolated images from the z-stack each channel were subjected to a circular normalization. C. Image representation of the point spread function in both the xy and zy directions for each channel.

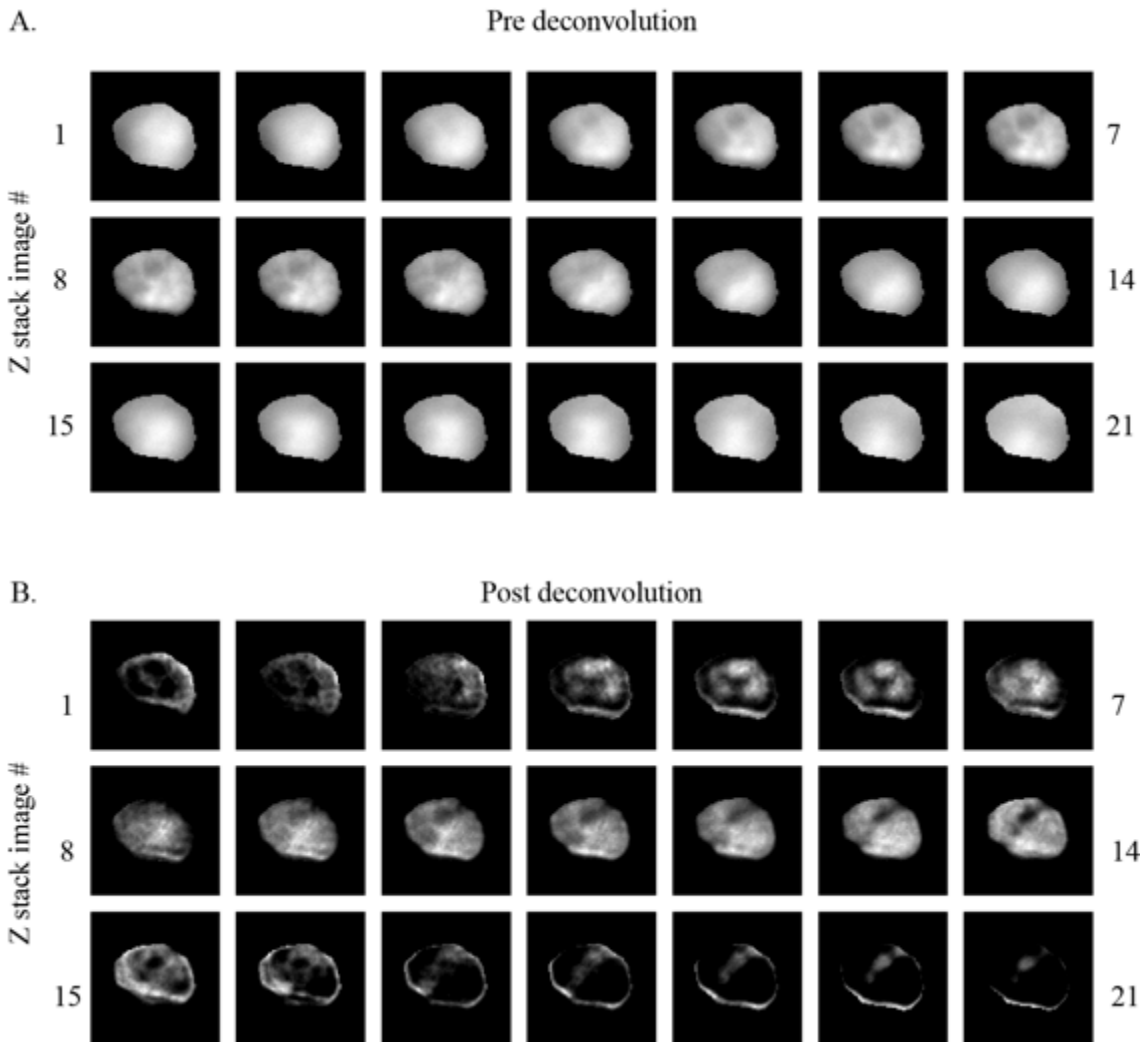


Figure 4.7: Deconvolution of the DAPI channel nuclei. A. A representative nucleus from each z-stack image prior to deconvolution. B. Nucleus from (A) with Richard-Lucy deconvolution using the DAPI point spread function.

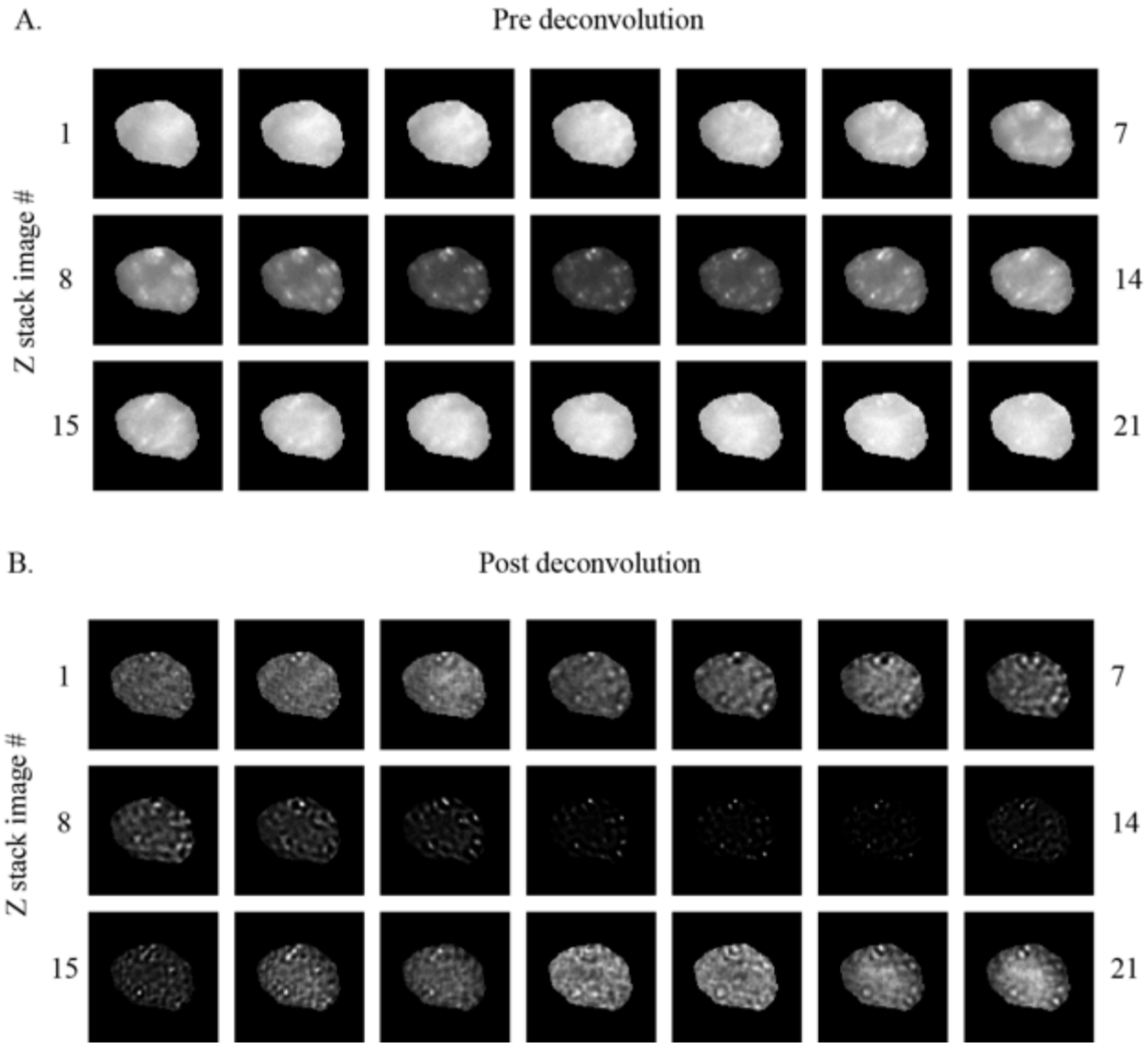


Figure 4.8: Deconvolution of the FITC channel nuclei. A. A representative nucleus from each z-stack image prior to deconvolution. B. Nucleus from (A) with Richard-Lucy deconvolution using the FITC point spread function.

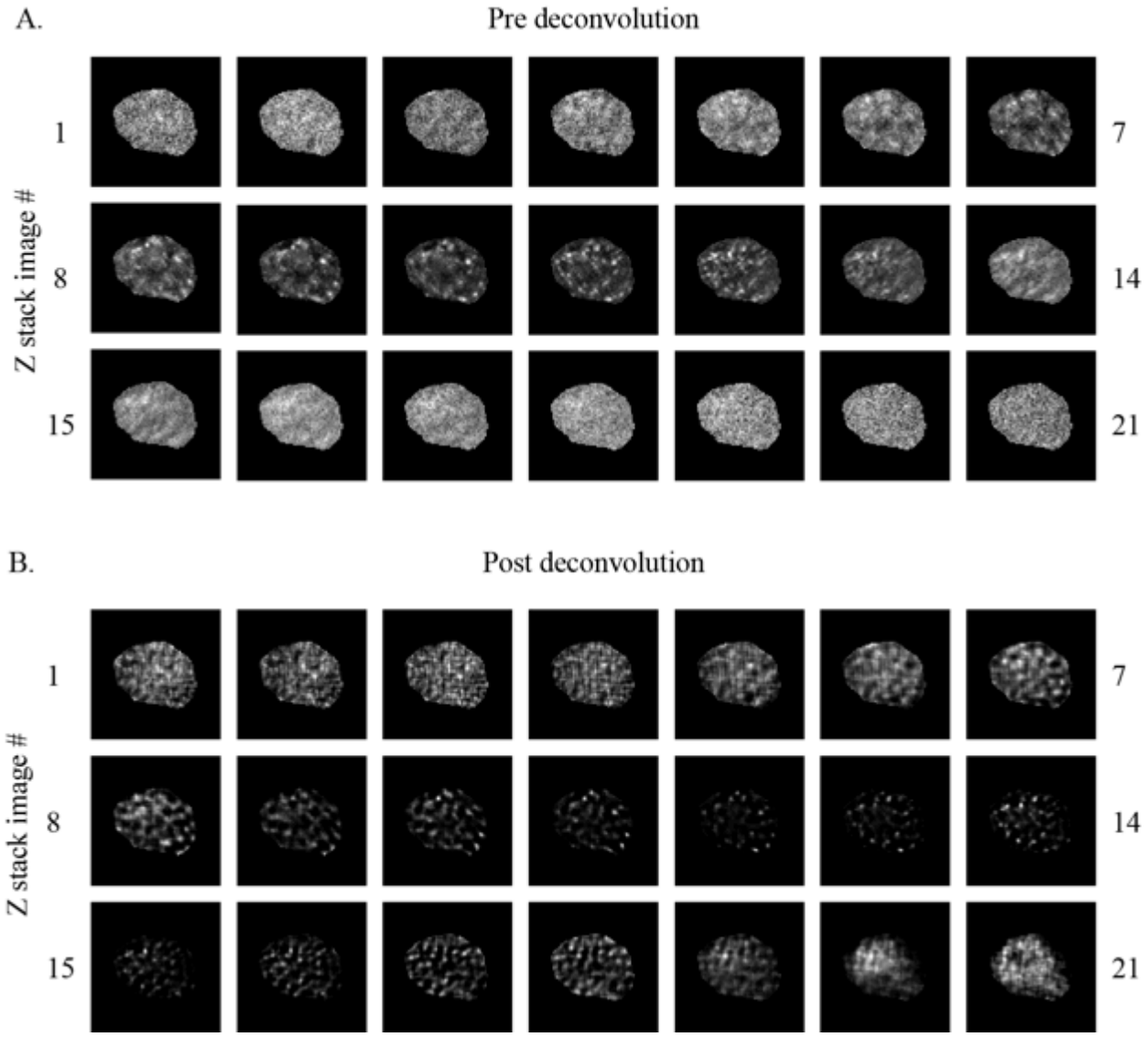


Figure 4.9: Deconvolution of the TRITC channel nuclei. A. A representative nucleus from each z-stack image prior to deconvolution. B. Nucleus from (A) with Richard-Lucy deconvolution using the TRITC point spread function.

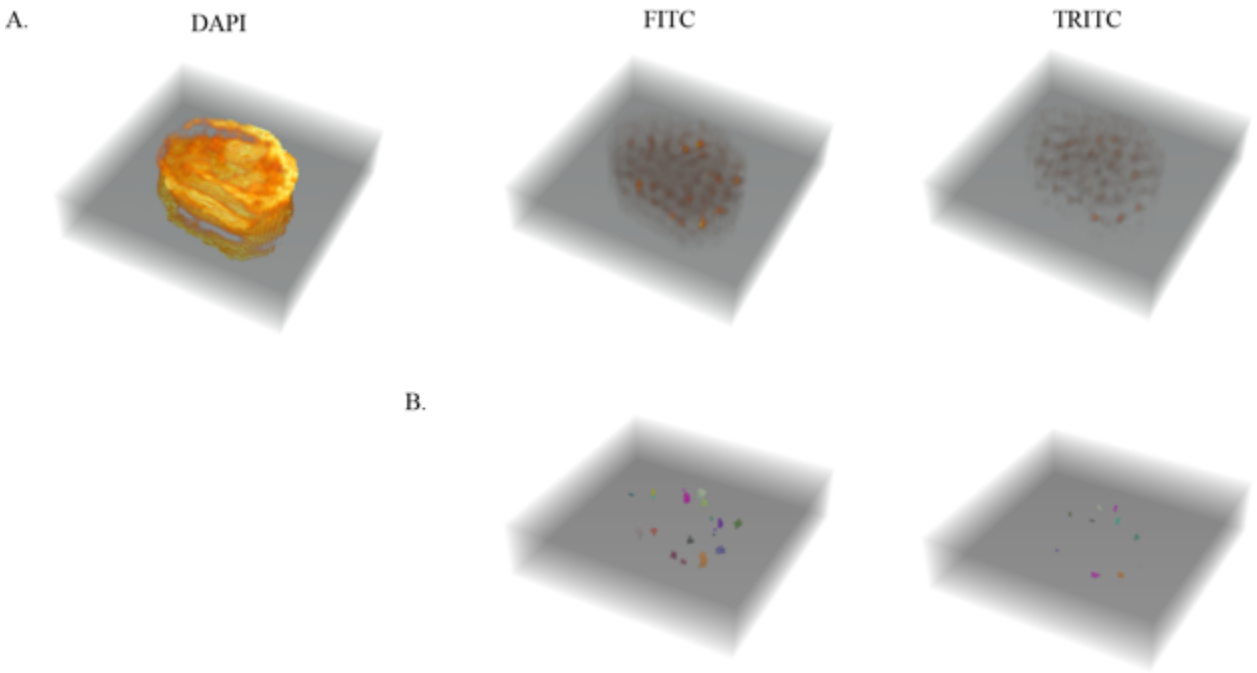


Figure 4.10: Three dimensional reconstructions. A. 3D reconstruction of the DAPI, FITC, and TRITC channels of a nuclei. B. Labeling of bright spots in the FITC and TRITC image.

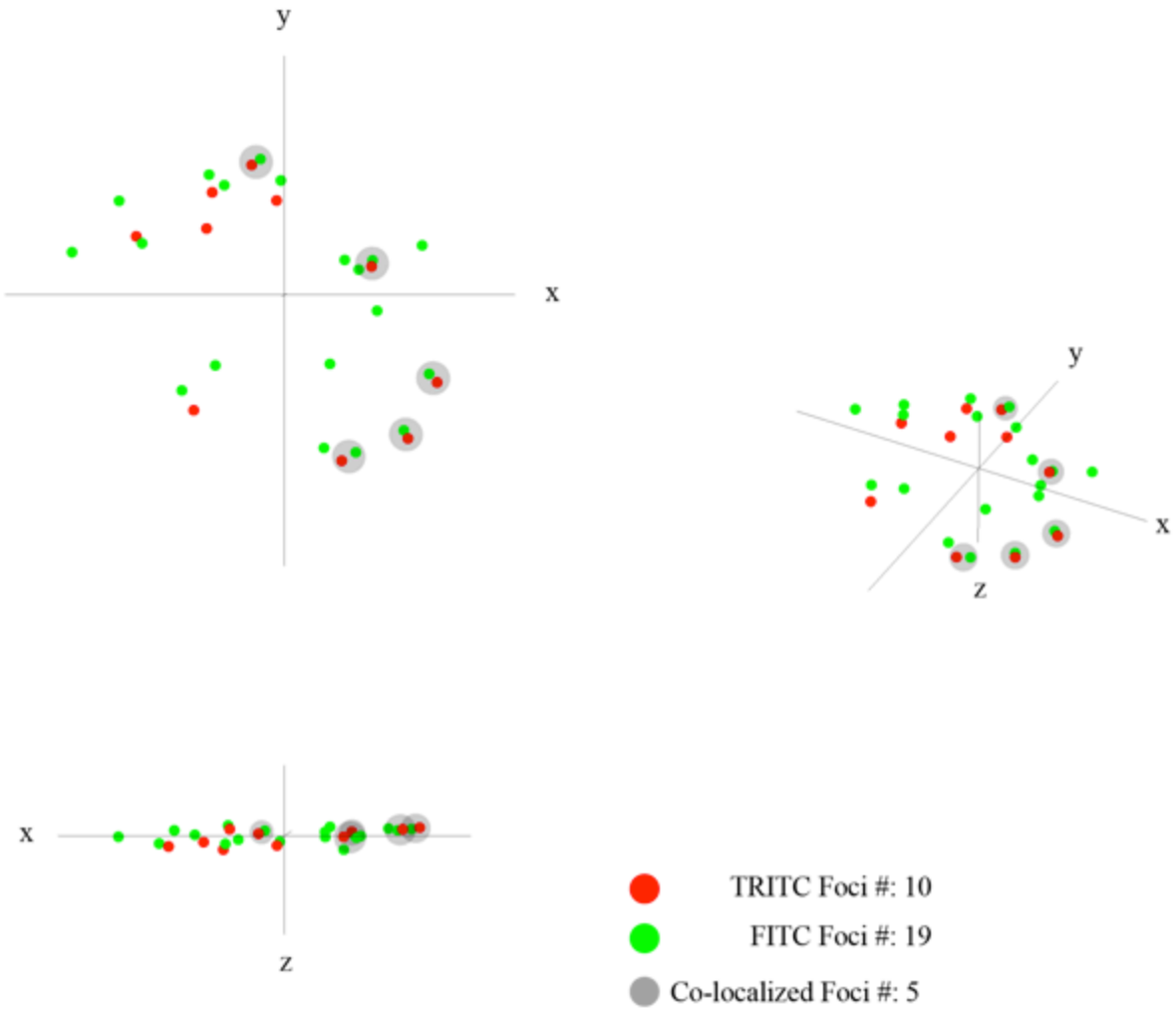


Figure 4.11: Quantification of Foci. Graphic representation of FITC (green dots) and TRITC (red dots) foci identified. Co localized foci are considered within two pixels of each other.

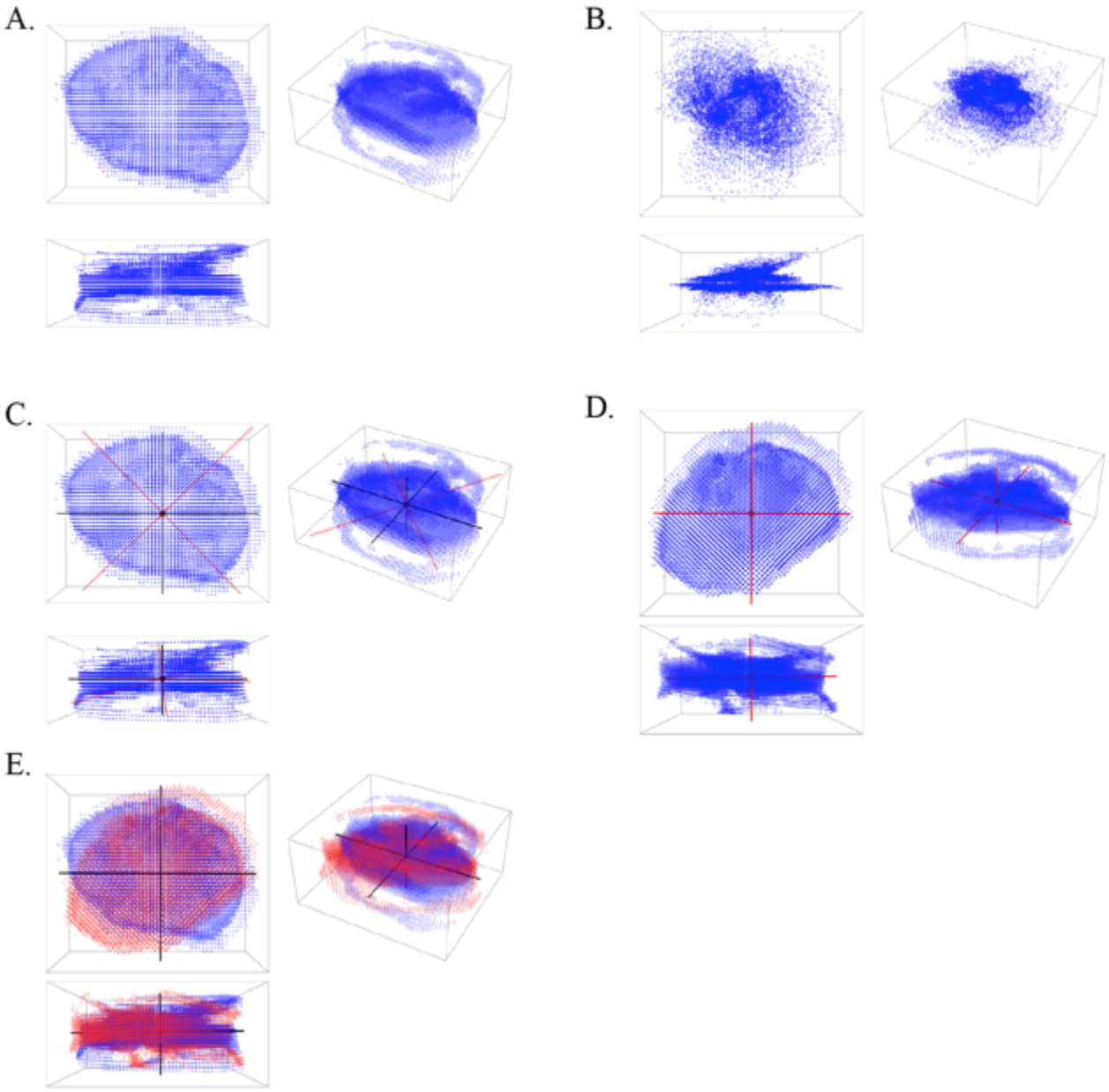


Figure 4.12: Normalization of nuclei orientation. A. 3D graphical plot of a nucleus. B. Mass distribution plot of nucleus in A. B. Red axes represent the moment of inertia (principle axes). D. The nuclei is rotated about the principle axes such that it is perpendicular to the x-axis. E. The input nucleus is show in blue and the rotated nucleus in red.

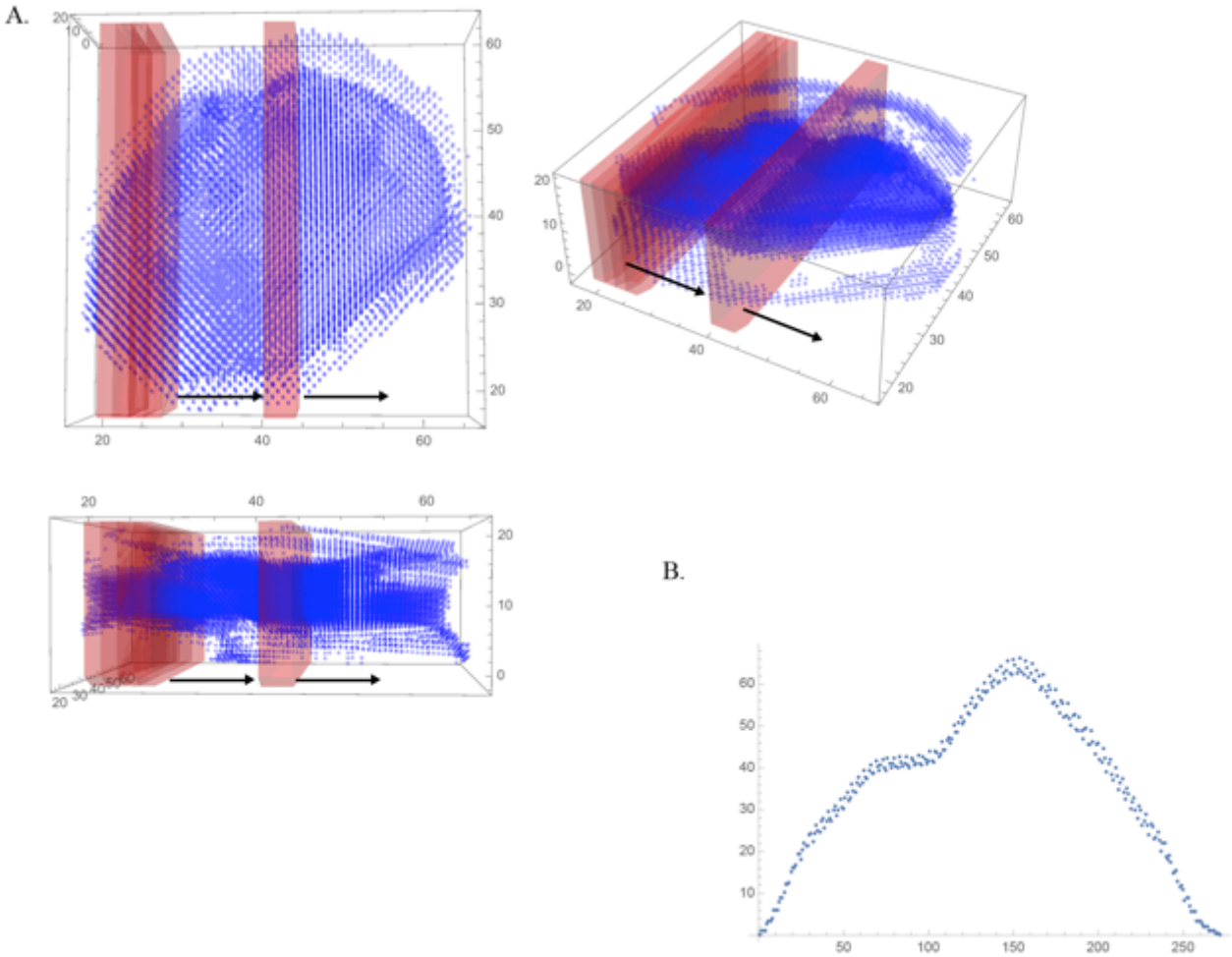


Figure 4.13: 3D volume scanning for intensity profile. A. 3D volume rendering showing a nucleus (Blue). The scanning window (Red) moves across the nuclei collecting intensity values. B. Integrated plot of the scanned intensity values.

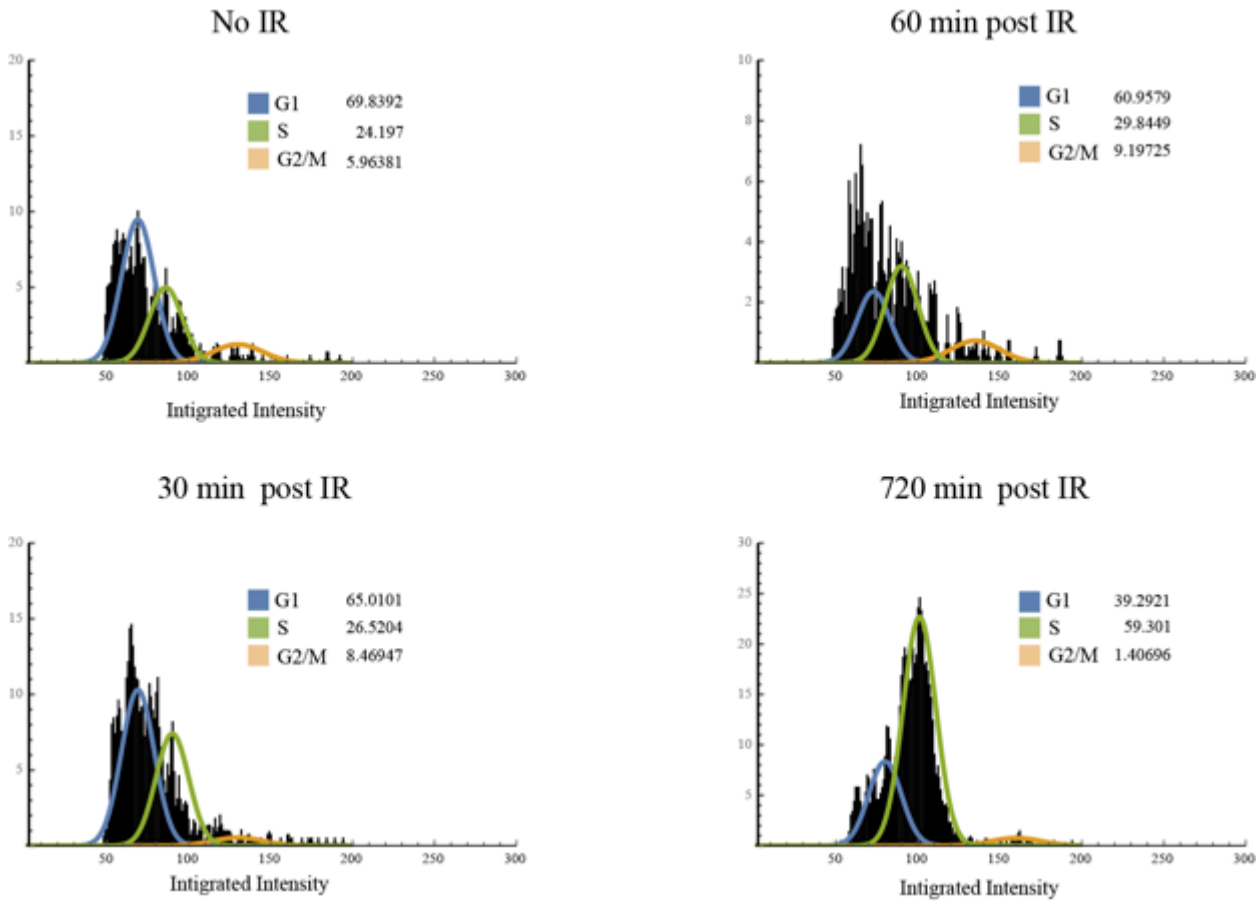


Figure 4.14: Cell cycle quantification using MANA. Cell cycle quantification of wild-type HEK293T cells analyzed 30, 60, and 720 minutes post exposure to 10 Gy ionizing radiation. The Watson method was used to identify phases of the cell cycle.

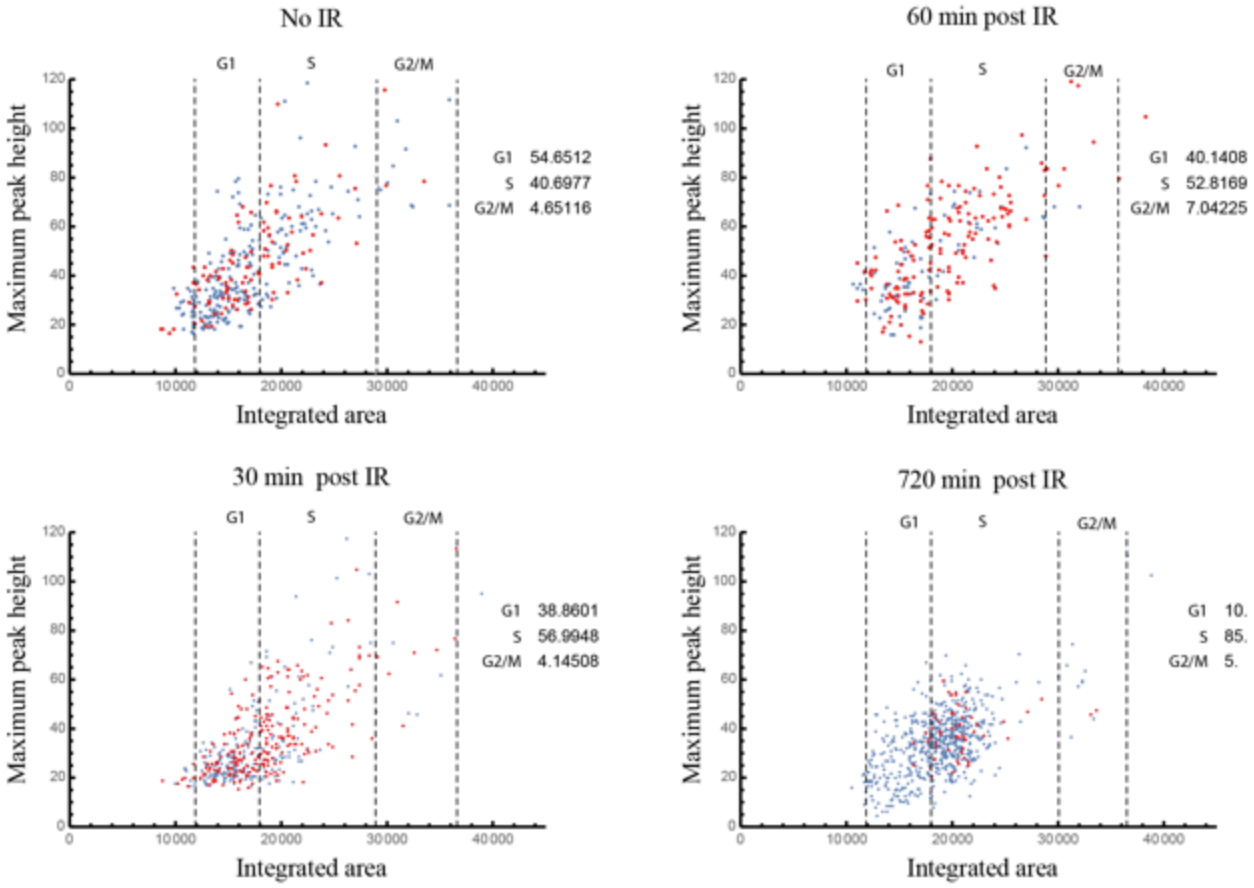


Figure 4.15: 53BP1 foci quantification and cell cycle staging. Scatter plots of wild type HEK293T cells co-stained with a DNA dye (Blue dots) and 53BP1 antibody (Red dots). The x-axis is the area under of the DNA dye intensity profile and the y-axis is the maximum intensity profile of the DNA stain. An example of an intensity profile for Hoechst staining is in Figure 4.13. The bounding regions for cell cycle phases were determined using the Watson method (Figure 5.14). Red dots represent nuclei that have greater than 10 53BP1 foci.

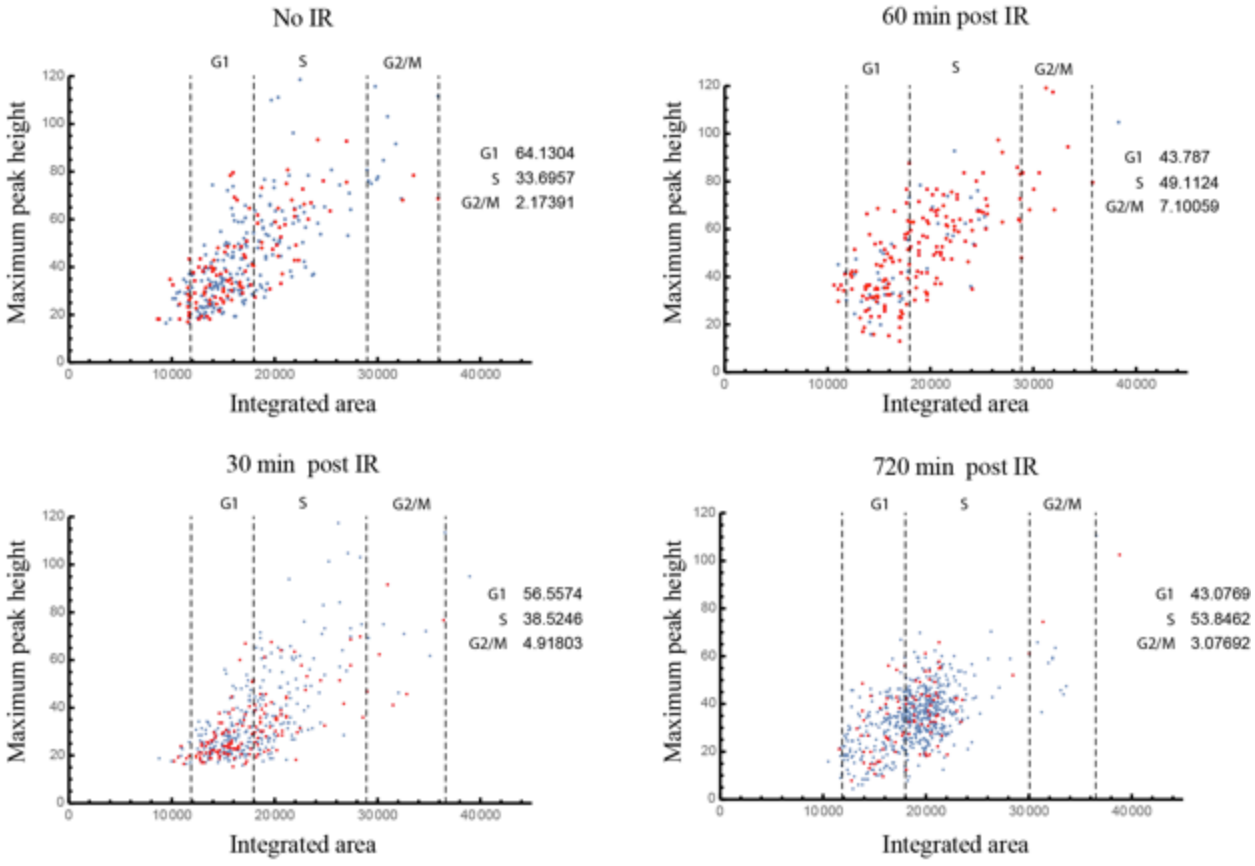


Figure 4.16: BRCA1 foci quantification and cell cycle staging. Scatter plots of wild type HEK293T cells co-stained with a DNA dye (Blue dots) and BRCA1 antibody (Red dots). The x-axis is the area under of the DNA dye intensity profile and the y-axis is the maximum intensity profile of the DNA stain. An example of an intensity profile for Hoechst staining is in Figure 4.13. The bounding regions for cell cycle phases were determined using the Watson method (Figure 5.14). Red dots represent nuclei that have greater than 5 BRCA1 foci.

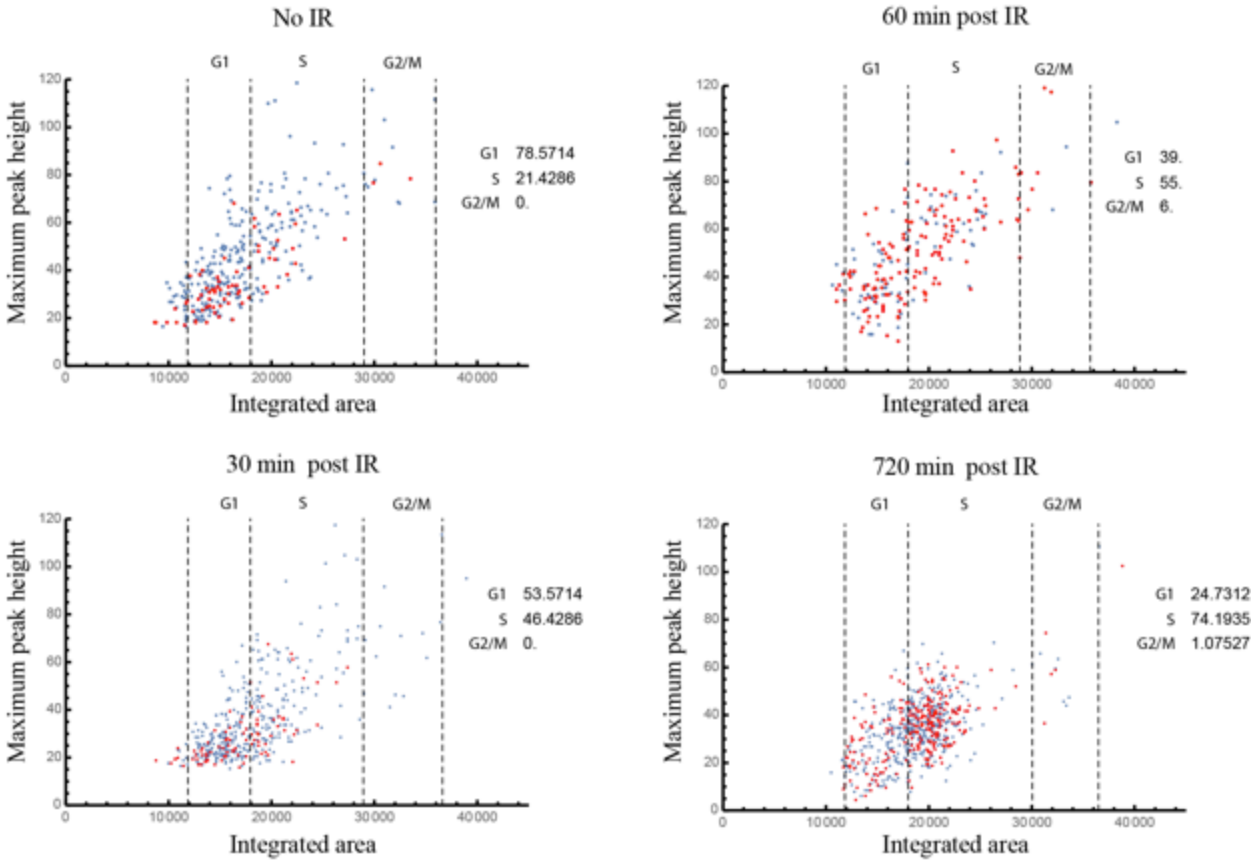


Figure 4.17: 53BP1/BRCA1 co-localization and cell cycle staging. Scatter plots of wild type HEK293T cells co-stained with a DNA dye (Blue dots) and 53BP1/BRCA1 antibodies (Red dots). The x-axis is the area under of the DNA dye intensity profile and the y-axis is the maximum intensity profile of the DNA stain. An example of an intensity profile for Hoechst staining is in Figure 4.13. The bounding regions for cell cycle phases were determined using the Watson method (Figure 5.14). Red dots represent nuclei that have greater than 5 53BP1/BRCA1 co-localized foci.

Chapter V: Histone Variant H2AZ Exchange at Sites Of Double Strand Breaks Requires Dot1l And H3k79 Methylation

Abstract

In the presence of DNA damage, the use of DNA repair pathways maintains genome integrity. In eukaryotic cells, the homologous recombination (HR) and non-homologous end joining (NHEJ) pathways are required for the repair of DNA double strand breaks (DSBs). The utilization of HR or NHEJ at the sites of DSBs is influenced by the local chromatin state. In mammals, histone modification and histone variant exchange at DSB sites generate an open chromatin state necessary for DSB repair to take place. How histone modifications contribute to histone variant exchange at DSB sites and how this processes contributes to DNA repair remains an important issue. In eukaryotes, phosphorylation of the histone variant H2A.X, termed γH2A.X, is an early histone modification that contributes to sensing DSBs. The propagation of the γH2A.X surrounding the break site leads to the recruitment of the p400/Tip60 containing NuA4 remolding complex via MDC1[137, 161]. p400 catalyzes the exchange of H2A for H2AZ variant leading to histone acetylation and ubiquitination at the DSB site[136]. Histone exchange and modifications at the DSB site transitions the chromatin into an open state that is permissive for HR repair. Here we show that Disruptor of telomeric siliencing-1 Like (DOT1L) is required for H2AZ exchange at DSB sites. DOT1L and its homologue in yeast, Dot1, are the only histone 3 lysine 79 (H3K79) methyltransferase in eukaryotic cells responsible for deposition of mono, di and tri methyl marks [101-103]. Cells from *Dot1L*$^{-/-}$ mice have increased genomic instability and defects in DNA repair. Loss of DOT1L and its methylation activity resulted in decreased H2AZ incorporation at DSB sites, increased amounts of single strand DNA (ssDNA) and decreased HR repair activity. We hypothesize that DOT1L and H3K79 methylation marks facilitate histone exchange at DSB sites generating a chromatin state that is permissive for HR repair. In the

absence of either DOT1L or H3K79 methylation the NHEJ repair pathway is utilized leading to genomic instability.

Material And Methods

Dot1L Knockout mice

Dot1L mutant mice were generated as described [162].

Ex vivo erythroid differentiation assay

Cells from E10.5 *Dot1L*$^{-/-}$ and wild-type yolk sacs cultured in M3434 methylcellulose medium (StemCell Technologies) containing the cytokines SCF, IL-3, IL-6, and EPO, which promote the growth of erythroid and myeloid progenitors.

Cell Culture and Cas9 Targeting

Embryonic day 10.5 (E10.5) embryos were obtained from timed matings between heterozygous Dot1L mutant mice, mouse embryonic fibroblasts (MEF) were isolated and cultured in DMEM, high glucose, L-glutamate, Sodium Pyruvate (Life Technologies, 11995-040), 10% fetal bovine serum (Life Technologies 10082-147), supplemented with 20 mM HEPES (Life Technologies 15630-080), L-glutamte, and Non-essential amino acids. HEK293T cells were cultured in DMEM, high glucose and 10% fetal bovine serum. *DOT1L*^{Stop} and *DOT1L*^{Y312A} cell lines were generated by Cas9/ssDNA directed cleavage/repair. PX330 plasmid (Zhang lab) was modified to express an eGFP-P2A-Cas9 gene (PXeC) and a gRNA targeting amino acid 312 of DOT1L. A 150 nt single stranded oligonucleotide, antisense to transcription, was used to direct repair of the Cas9 cleaved locus by destroying the PAM, adding a DdeI restriction site and either adding three tandem stop codons or nucleotides converting Tyrosine 312 to Alanine. HEK293T cells were transfected with a 1:1 molar ratio of PXeC and ssDNA using Effectine (Qaigen) in a six well dish. Cells were sorted for eGFP expression 48 hrs after

transfection (FACSaria II, BD). Colonies were picked 5 days later, genomic DNA was extracted and a PCR/restriction digest with DdeI was used to initially identify DOT1L KO and Y312A positive clones. Western blot and Sanger sequencing was used to verify clones.

Immunofluorescence

E10.5 MEF were plated into 24 well dishes containing L-Poly lysine (75 000 -150 000 MW) treated round cover slips. For end resection assay Cells were grown in conditioned media (1:2) for 48 hrs prior to Brdu addition. Cells were cultured in 10μM BrdU for 48hrs and exposed to ionizing radiation. Cells were fixed at 15 min, 30min, 60min, 120min and 240min post irradiation with 4% Paraformaldhyde for 30 min at room temperature, cells were washed with 200mM Glycine/1XDPBS for 10 min, permeabilized with 0.2%TritonX100/1XDPBS for 30min, blocked with 1%BSA/1XDPBS for 60 min. For BrdU incorporation positive control cells were treated after Glycine wash with ice cold 1N HCl for 20min, 2N HCl for 10 min, neutralized with room temperature phosphate buffer, pH7.4 for 10 min, flowed with permeablization, and blocking as above. BrdU was visualized by staining with primary antibody MoBu-1 (1:500) for 1 hr, followed by secondary antibody AlexFlour 588 (1:1000) for 2 hrs.

Western Antibodies and immunefluorescence

H3K79 di meth (ab3594), H3K79 mono meth (ab2886), H3 (ab1791), RAD51 (ab63801), 53BP1(ab36823), BRCA1,(ab16780), H2AZ (#2718, CellSignaling), H4K acetylation (ab7311).

qPCR and DNA damage pathway primers

qPCR was performed using Power SYBR® Green PCR Master Mix (#4367659, Thermofisher Scientific), primers with annealing temperature of 61^{o}C and 40 cycles on ABI 7900 system. Two-tailed Student's t-test was used for significance testing. Primers for mouse DNA damage pathway genes were designed using IDT PrimerQuest Tool. Total RNA was

isolated from cells of E10.5 *Dot1L*$^{-/-}$ and wild-type mice that were 6 hours post ionizing radiation treatment (2 Gy).

Gene	**Fwd**	Rev
Anxa5	CGAATAGAGACCCTGATACTGC	ACTGCGTGTCCCAAAGATG
Atm	GATCCTTCCCACTCCAGAAAC	ACTCCGCATAACTTCCATCG
Atr	TCCTGAAAATAAGTCCCCATCTG	TGGCAGAAACAAGGTCACATC
Atrip	CTTCTCCTGAGGCTGTACATG	ACAGTCAGAACCAATGCCAG
Bard1	TCCCGCTGTGCTAATATTCTG	GTCGGTTTATCTTGAGGTCTAGG
Blm	TGGATCAGAAAGCATCACCC	GTAAAGTGTCAGCCATTGTGTC
Brca1	AGGGAAGCACAAGGTTTAGTC	CCTCATTCAAACGCTCACAAG
Brca2	CTTTAGAGCCGGTTACCTGTG	AGGACTGCTTGGAGACTTTTC
Cdk1	TGCAGGACTACAAGAACACC	GCCATTTTGCCAGAGATTCG
Cdk2	GCATTCCTCTTCCCCTCATC	GGACCCCTCTGCATTGATAAG
Chd1l	TCTCTGTCCCTTGTCTGTTTTG	ACGTGAAAACCTGACTCCTG
Chek1	ACTGGTTGACTTCCGACTTTC	ACAGCTTCCTCATGTAACAGG
Chek2	CTGAGGACCAAGAACCTGAAG	CCATCGAAGCAATATTCACAGC
Dclre1c	TCGTGTGGCTGAACTGTAAAG	TTCTGTCCGTTGTGAGATGG
Eme1	TGGTAGAGGATGGAGATGACTG	AGTTACAAAGCTCCGAAGGG
Ercc1	CCCGTGTTGAAGTTTGTGC	CCGTTCATGGATGTAGTCTGG
Ercc4	CAGACGGCCATACTAGACATC	CCACAAAGGGTCCAAGTAATG
Exo1	CAGTAGGGATTCAGGTTCAGAG	ACTAGACCTACAGAGCCCAG
Gen1	TGCCGTTGATTTGAGTCTCTG	TCCCCTTCCATTACAAACACC
H2afx	AAGTCGCGCTCTTCACG	TCTCGGCAGTGAGGTACTC
H2afz	GTATCACCCCTCGTCACTTG	TCTTCTGTTGTCCTTTCTTCCC
Herc2	ACACCAGGCATACTTTCGG	GTTCAGGCTTTTCACCATTCG
Kat5	CACATCGTGGGCTACTTCTC	GGTCTTCCCTTCTACTTTCGAG
Lig3	AGTGGGAATGAAGAGGAAAGC	GTGTAGAAGGTGGCAAGTAGAG
Lig4	CCCCTGTGATTGCTGACTTAG	TCTCGTGGCTTCAATTCTGG
Mdc1	TGATTGACTGGGATGCTGAAG	CATTCTTGCCAAGGTAGAGGG
Mre11a	CAGAAAGGAAGCACAAAACTCG	TCACAAACATCCGATAGAGCC
Mus81	GATATTAAGGAGTCGGCTGGC	GAGAGGGTTTGTGGAAGGC
Nbn	CCTTTCCCAAGTCCTGTCAG	CAGCCCAGGGATCTTTACTTC
Nhej1	GTTATGCCTTGCTGATTTCGG	ATCCAGGTGACAGAGCAAAG
Palb2	AACCCAGGAAACCTACAGTTG	AGCAGTTACGATACATGGCTC
Parg	GATTGTTTCACGGCTGTTCAC	ATCGTCCTTTTCACTCCCATC
Parp1	CAGCGAGAGTATTCCCAAGG	AAGCCATCCACCTCAACATC
Pias4	CCTTACCCACCTCATGTATCTG	TCACCCCAATCGTCTTCAAC
Prkdc	TTCTCCATACCCAAACCCAAG	CTCCCAGTCAGCCAATCAAA
Rad50	AGTGTGCCTGACAGATGTTAC	GCTTTTCCTGCTTTTCCTGG
Rad51	TGGTGTCGCAGTGGTAATC	CTCCCCTCTTCCTTTTCTCAG

Rad52	CCAGCATGTTCTAGGTAGCAG	TCCTTTTGTGACGGTCTAGTG
Rbbp8	CCAAGCAACCAAGATACGTCC	TCCATGTCCACTGTTTCTCC
Rmi1	GATCCTACAGTTCCAGTCATTCC	GCCCCAAAACAGCATCTAAG
Rmi2	TCCAGAAAATCAGATCCAGTCAC	CAGATGTCTTCCTCAAGCTCAG
Rnf168	TGAACGACTTGCTGGAGAAC	TTCGACAAATGGGACACTCTAC
Rnf8	ACGTTCAGAATCAGGGTCAAG	GGTACTTGCTCCCATCACATG
Rpa1	GTCTTGTCCTTTATCAGTTGTGC	CAATTTCCGAGATGCCAGTTTC
Slx1b	GGAAGAGGATTTGGAGTTAGAGG	GAGAAGCAGGAAAGGTACAGG
Slx4	CTGAACTACCTCTACATGGCAG	TCTGGCTGTTCTCCCTCTAAG
Top3a	GTCTTGTCCTTTATCAGTTGTGC	CAATTTCCGAGATGCCAGTTTC
Trp53	ATGTTCCGGGAGCTGAATG	CCCCACTTTCTTGACCATTG
Trp53bp1	CCCTGATGCTTTCCGATCTAC	TCTGTCTCCATTGCTTCGTC
Ube2i	GAACCAAATATTCAAGACCCAGC	CAATCCCTTCCTCGTCATGG
Ube2n	TGATGTAGCCGAGCAATGG	TTGGCAGAACAGGAGAAGTG
Uimc1	TGATGGAGCAGGAAACAGTG	GTGGAACCAGGGACTTACATAG
Usp28	GTGACATTACCACTTCCCTCG	CATTCATCCCCTCAGAGCTG
Xrcc1	GCTGGGACCGTGTTAAAATTG	GTCACTGTCACCTTCTGAGATG
Xrcc4	TGGGACAGAACCTAAAATGGC	GGGTAGTGAAGAGGCAAGC
Xrcc5	TGACTGCTCAGGACGTTTTC	CCTTGGTGATGTTCCCTTCTG
Xrcc6	CGGGAAACAAATGAACCAGTG	TCCTCTTTCTCCAGCACAATC

Table 5.1: qPCR primers for DNA damage repair pathway genes in mice.

ssDNA assay

ssDNA was measured using a method designed to enrich for ssDNA and sequenced on an Oxford Nanopore MinION. The methodology is described in Chapter III.

BEL-Chip

BEL-ChIP was performed on wild-type, *DOT1L*^{Stop} and *DOT1L*^{Y312A} cell lines using H2AZ (#2718, CellSignaling), H4K acetylation (#ab7311, Abcam) antibodies. BEL-ChIP is described in Chapter II.

Cell cycle staging and foci analysis

Cell cycle staging and foci quantification were performed on cells isolated from E10.5 *Dot1L*$^{-/-}$ and wild-type mice using MANA. Cell cycle staging and foci quantification were

performed on wild-type, *DOT1L*^{Stop} and *DOT1L*^{Y312A} cell lines using MANA. MANA is described in Chapter IV.

HR/NHEJ assays

Wild-type, *DOT1L*^{Stop} and *DOT1L*^{Y312A} HEK293T cell lines expressing an HR reporter [163], alt-NHEJ reporter [69], and a NHEJ reporter [69], were transfected individually with Effectine (Qiagen) and selected for using puromycin. Five days post-transfection, cells were transfected with the I-SceI plasmid using Effectine (Qiagen), and GFP-positive cells were counted 48 hr later using a Invitrogen Attune.

Nucleosome stability assay

The nucleosome stability is described in [144]. Wild-type, *DOT1L*^{Stop} and *DOT1L*^{Y312A} cell lines were exposed to 10 Gy ionizing radiation and harvested at 30, 60, and 240 minutes post IR. Cells were washed with ice-cold PBS and re-suspended in 500 µL of Stability buffer (20 mM Hepes, pH 7.9, 0.5 mM DTT, 1.5 mM MgCl 2, 0.1% Triton) containing 1.0 M NaCl, Roche Protease Inhibitor cocktail) and agitated constantly for 40 min at 4 °C. Cells were collected by centrifugation at 100,000 × g (Beckmann Ultracentrifuge) for 20 min, and the supernatant was analyzed by western blot.

Comet assay and analysis

The neutral comet assay was performed using the Trevigen Comet Assay kit (Trevigen, Gaithersburg, MD). Cells were isolated from yolk sacs of E10.5 *Dot1L*$^{-/-}$ and wild-type mice grown *ex vivo* differentiation media (see above), after 4 days cells were harvested and resuspended in cold PBS, and an aliquot of cells (1000/10 µl) was added to 100 µl of molten LMA agarose and spread onto a comet slide. The slide was incubated at 4 °C for 10 min to and transferred to cold lysis solution for 60 min at 4 °C. A denaturation step was performed in 50

mM Tris base, 150 mM pH 9, for 30 min at 4 °C. The slides were then subjected to electrophoresis with cold TAE buffer, pH 8.2 at 25 V for 30 min at 4 °C, and immersed in DNA precipitation solution (100 mM NH4Ac in 95% ethanol) and then in 100% ethanol for 30 min and air dried. DNA was stained with 100 µl SYBR Gold (Invitrogen, 1:30,000) for 20 min and immediately rinsed with dH2O and air-dried. The slides were imaged using an upright Nikon Eclipse 80i fluorescence microscope at ×20, and analyzed using CometScore Pro (TriTek Corp).

For the repair of induced DNA damage assay cells were isolated from E10.5 *Dot1L*$^{-/-}$ and wild-type mice and exposed to 2 Gy of ionizing radiation and a comet assay was performed as above at 24, 72, and 96 hours post exposure. Comets were analyzed using a custom comet analysis software development in Mathematic programing language; the code for CometTool is provided below.

What is this figure? Figure legend?

```
CreatePalette[
 DynamicModule[{ddir = "", fdir = "", idir = "", x = "", y = "", z = "", a = "",
   dimg = Image[{{1}}], fimg = Image[{{1}}], data = "", locapts = "", updatedpts = "",
   mask = Image[{{1}}], fullfimg = Image[{{1}}], fsimg = Image[{{1}}], simg = "",
   c = "", f = {}, k = "", o = "", , j = {1, 2}, rimg = "", report = "", r = "",
   headcenter = "", head = "", headdata = "", tailkeys = "", tailvalue = "",
   tailass = "", selecttailkeys = "", tailimg = "", taildata = "", taillength = "",
   headcoords = "", headint = "", tailcoords = "", tailint = "", totaltailint = "",
   totalheadint = "", headkeys = "", headass = "", selectheadkeys = "", headvalue = "",
   l = "", back = "", tailnorm = "", headnorm = "", tailcm = "", headcm = "",
   ptaildna = "", m = "", v = "", dist = "", tdist = "", otm = "", tlen = "", tcdir = ""},
  Manipulate[
   If[x === "", dimg = Image[{{1}}],
    dimg = Import[StringJoin[ToString[ddir], ToString[x]]]];

   locapts = {};
   If[x === "", updatedpts = "", updatedpts =
     Dynamic[{{#[[1]], #[[2]]}, {#[[1]] + xlength, #[[2]] + ylength}} & /@ locapts]];
   If[a === "", a = 0, a = Length[data[[All, 1]]]];

   If[x === "", Show[Image[{{1}}]],
    Show[dimg, Graphics[
      {White, MapIndexed[Text[#2[[1]], {(#[[1]] + xlength/2), (#[[2]] + ylength/2)},
            BaseStyle → Large] &, #[[All, 1]]] & /@ (updatedpts), EdgeForm[
        Directive[Dashed, Red]], Opacity[0], Rectangle @@@ # & /@ (updatedpts)}]]],
```

```
    {locapts, Locator, LocatorAutoCreate → True, Appearance → Style["·", 50, White]},

    OpenerView[{"Comet Selection Settings", Column[{
        Button["Select Comet Directory", ddir = SystemDialogInput["Directory"],
         ImageSize → {200, 25}, Method -> "Queued"],
        Dynamic[PopupMenu[Dynamic[x], Last[StringSplit[#, "/"]] & /@
           FileNames["*.tif", ddir]]],
        OpenerView[{"Comet Bounding Box", Column[{
            Control[{{xlength, 90, "Bounding Box X length"}, 1, 300, 5}],
            Control[{{ylength, 86, "Bounding Box Y length"}, 1, 300, 5}]
           }]}],

        Button["Select Ruler Image", fdir = SystemDialogInput["Directory"],
         ImageSize → {200, 25}, Method -> "Queued"],
        Dynamic[PopupMenu[Dynamic[y], Last[StringSplit[#, "/"]] & /@
           FileNames["*.tif", fdir]]],
        Button["Convert Pixel to µm", {m = MorphologicalComponents[ColorNegate[

             Binarize[Import[StringJoin[ToString[fdir], ToString[y]]]]]],
v = Values[ComponentMeasurements[m, "Centroid"]],
         dist = Sort[v[[All, 1]]][[#]][[2]] - Sort[v[[All, 1]]][[#]][[1]] & /@
           Flatten[Transpose[{Drop[Range[Length[v]], -1],
                Drop[Range[Length[v]], 1]}] & /@ Range[1], 1],
         tdist = (Total[dist] / Length[dist]) / 10,
         Clear[m]},
        ImageSize → {200, 25},
        Method -> "Queued"],
        Panel[Dynamic[If[tdist === "", Image[{{1}}], tdist], SynchronousUpdating →
           False, ContinuousAction → False], "Pixel per µm", Left]
       }]}],


    Button["Isolate Comets", {idir = SystemDialogInput["Directory"],
      tcdir = CreateDirectory[
        StringJoin[ToString[idir], ToString[First[StringSplit[x, {"/", "."}]]]]],
      rimg = ImageTrim @@ {ColorConvert[Import[StringJoin[ToString[ddir],
                ToString[x]]], "Grayscale"], #} & /@ # & @@ (updatedpts),
      Table[Export[FileNameJoin[{ToString[tcdir], StringJoin["C_",
           ToString[i], ".tif"]}], rimg[[i]], "TIFF"], {i, Length[rimg]}],
      Clear[rimg]},
     ImageSize → {200, 25},
     Method → "Queued"],
    Panel[Dynamic[PopupMenu[Dynamic[z],
       If[rimg === "", Range[Length[FileNames["C*.tif", tcdir]]],
        Range[Length[FileNames["C*.tif", tcdir]]]]], "Comet Selector", Left],

    OpenerView[{"Delete Isolated Images",
      Column[{ Button["Delete Isolated Images", {DeleteFile[#] & /@ FileNames[
           "*.tif", tcdir], z = ""}, ImageSize → {200, 25}, Method → "Queued"]}]}],
```

```
OpenerView[{"Measure Comet", Column[{
    Button["Measure Comet", k = 1, ImageSize → {200, 25}, Method → "Queued"],
    Control[{{radius, 13, "Head Radius"}, 1, 30, 0.5}],
    Control[{{bxxlength, 20, "Comet Length"}, 1, 200, 1}],
    Control[{{sbxxlength, 0, "Shift Uppder Background box"}, 0, 200, 1}],
    Control[{{sbyylength, 0, "Shift Lower Background box"}, 0, 200, 1}],
    TogglerBar[Dynamic[j],
     {1 → "upper background box", 2 → "lower background box"}],
    Pane[DynamicModule[{pt = "", img = Image[{{1}}]},
      Column[{Panel[Dynamic[LocatorPane[

            If[z === "", img = Image[{{1}}], img = Import[FileNameJoin[
                {ToString[tcdir], StringJoin["C_", ToString[z], ".tif"]}]]];
            If[k === 1,
             headcenter = pt;
             Which[Total[{1, 2}[[j]]] === 3,
              back = backgroundsmooth[
                img, 0, headcenter, radius, bxxlength, sbxxlength],
              Total[{1, 2}[[j]]] === 1, back = backgroundsmoothupper[
                img, 0, headcenter, radius, bxxlength, sbxxlength],
              Total[{1, 2}[[j]]] === 2, back = backgroundsmoothlower[
                img, 0, headcenter, radius, bxxlength, sbxxlength]];
             tailnorm = tailplotdata[img, 0, headcenter, radius,
               bxxlength, sbxxlength, sbyylength, back];
             headnorm = headplotdata[img, 0, headcenter, radius, back];
             tailcm =
              Total[#[[1]] * #[[2]] & /@ tailnorm] / Total[tailnorm[[All, 2]]];
             headcm = Total[#[[1]] * #[[2]] & /@ headnorm] /
               Total[headnorm[[All, 2]]];
             ptaildna = (Total[tailnorm[[All, 2]]] / (Total[tailnorm[[
                      All, 2]]] + Total[headnorm[[All, 2]]])) * 100;
             otm = (headcm - tailcm) / tdist;
             tlen = Length[tailnorm] / tdist;
             k = k - 1;];
            Dynamic[pt],
            If[pt === "", Show[heatmap[img, 0]], Which[
              Total[{1, 2}[[j]]] === 3, Show[heatmap[img, 0],
               Graphics[{EdgeForm[Directive[Thick, Red]], Opacity[0],
                 Disk[pt, radius], EdgeForm[Directive[Thick, Blue]], Opacity[0],
                 Rectangle[{pt[[1]] + radius, pt[[2]] + radius * 2 + sbxxlength},
                  {pt[[1]] - radius - bxxlength, pt[[2]] + radius + sbxxlength}],
                 EdgeForm[Directive[Thick, Black]], Opacity[0],
                 Rectangle[{pt[[1]] - radius, pt[[2]] - radius - sbyylength},
                  {pt[[1]] - radius - bxxlength, pt[[2]] + radius + sbxxlength}],
                 EdgeForm[Directive[Thick, Blue]], Opacity[0],
                 Rectangle[{pt[[1]] - radius - bxxlength,
                   pt[[2]] - radius - sbyylength}, {pt[[1]] + radius, pt[[2]] -
                    radius * 2 - sbyylength}]}]], Total[{1, 2}[[j]]] === 1,
              Show[heatmap[img, 0], Graphics[{EdgeForm[Directive[
                    Thick, Red]], Opacity[0], Disk[pt, radius],
                 EdgeForm[Directive[Thick, Blue]], Opacity[0],
                 Rectangle[{pt[[1]] + radius, pt[[2]] + radius * 2 + sbxxlength},
                  {pt[[1]] - radius - bxxlength, pt[[2]] + radius + sbxxlength}],
                 EdgeForm[Directive[Thick, Black]], Opacity[0],
```

```
                    Rectangle[{pt[[1]] - radius, pt[[2]] - radius - sbyylength},
                     {pt[[1]] - radius - bxxlength, pt[[2]] + radius +
                       sbxxlength}]}]], Total[{1, 2}[[j]]] === 2,
                 Show[heatmap[img, 0], Graphics[{EdgeForm[Directive[
                      Thick, Red]], Opacity[0], Disk[pt, radius],
                    EdgeForm[Directive[Thick, Black]], Opacity[0],
                    Rectangle[{pt[[1]] - radius, pt[[2]] - radius - sbyylength},
                     {pt[[1]] - radius - bxxlength, pt[[2]] + radius + sbxxlength}],
                    EdgeForm[Directive[Thick, Blue]], Opacity[0], Rectangle[
                     {pt[[1]] - radius - bxxlength, pt[[2]] - radius - sbyylength},
                     {pt[[1]] + radius, pt[[2]] - radius * 2 - sbyylength}]}]]],
               {{0, 0}, {ImageDimensions[img][[1]], ImageDimensions[img][[2]]}},
               Locator, LocatorAutoCreate → True, Appearance → Style["·", 40, Black]
              ], SynchronousUpdating → False, ContinuousAction → False]]}],
        InheritScope → True], Scrollbars → True, ImageSize → {400, 350}],
      Row[{Panel[Dynamic[otm, SynchronousUpdating → False,
          ContinuousAction → False], "Olive Tail Moment (µm)", Left],
        Panel[Dynamic[ptaildna, SynchronousUpdating → False,
          ContinuousAction → False], "% Tail DNA", Left],
        Panel[Dynamic[tlen, SynchronousUpdating → False,
          ContinuousAction → False], "Tail Length (µm)", Left]}]}]}],

Button["Collect Comet Data",
 f = Catenate[{f, {<|"File Location" -> FileNameJoin[{ToString[x], StringJoin[
            "C_", ToString[z]]}], "File Name" → StringTemplate["C_`1`"][z],
       "Image" → Import[FileNameJoin[{ToString[tcdir], StringJoin["C_",
            ToString[z], ".tif"]}]], "Olive Tail Moment (µm)" → otm,
       "%DNA in Tail" → ptaildna, "Tail Length (µm)" → tlen|>}}],
 ImageSize → {200, 25}, Method → "Queued"],
Panel[Dynamic[If[Length[f] ≤ 0, Image[{{1}}], Values[Last[f]][[1]]]],
 "Last Analyzed Image", Left],
Button["Delete Last Collected Comet", f = Drop[f, -1],
 ImageSize → {200, 40}, Method → "Queued"],

OpenerView[{"Data Reporting Settings", Column[{
    Button["View Current Comet Data",
     genreport[x, f], ImageSize → {200, 25}, Method → "Queued"],
    Button["Export Comet Data to CSV Report", reportsave[f, x],
     ImageSize → {200, 50}, Method → "Queued"],
    Button["Clear Current Data", Clear[y, f],
     ImageSize → {200, 25}, Method → "Queued" ]
   }]}],
```

```
ContentSize → {1024, 768},
ControlPlacement → Left,
ContinuousAction → False,


Initialization :> (

  genreport[fname_, countlist_] :=
   CreateDocument[{
     TextCell[TemplateApply[
       StringTemplate["Comet Report for `1` sample `2`, `4` post `3`"],
       StringSplit[fname, {"_", "+"}][[{1, 2, 4, 5}]]], "Title"],
     TextCell["Summary Results", "Section"],

     CellGroup[{
       Row[{ExpressionCell[
           BoxWhiskerChart[Dataset[countlist][[All, "Olive Tail Moment (µm)"]],
            PlotLabel → "Olive Tail Moment (µm)"], "Output"],
          ExpressionCell[BoxWhiskerChart[Dataset[countlist][[All,
             "%DNA in Tail"]], PlotLabel → "%DNA in Tail"], "Output"],
          ExpressionCell[BoxWhiskerChart[Dataset[countlist][[
             All, "Tail Length (µm)"]],
            PlotLabel → "Tail Length (µm)"], "Output"]}]}],

     CellGroup[{
       Row[{ExpressionCell[ListPlot[
            {Transpose[{Normal[Dataset[countlist][[All, "Tail Length (µm)"]]],
               Normal[Dataset[countlist][[All, "Olive Tail Moment (µm)"]]]}]},
            PlotLabel → "Tail Length (µm) vs. Olive Tail Moment (µm)",
            AxesLabel → {"Tail Length (µm)",
              " Olive Tail Moment (µm)"}], "Output"],
          ExpressionCell[ListPlot[{Transpose[{Normal[
                Dataset[countlist][[All, "Tail Length (µm)"]]],
               Normal[Dataset[countlist][[All, "%DNA in Tail"]]]}]},
            PlotLabel → "Tail Length (µm) vs. %DNA in Tail",
            AxesLabel → {"Tail Length (µm)" , " %DNA in Tail"}], "Output"],
          ExpressionCell[ListPlot[{Transpose[{Normal[
                Dataset[countlist][[All, "Olive Tail Moment (µm)"]]],
               Normal[Dataset[countlist][[All, "%DNA in Tail"]]]}]},
            PlotLabel → "Olive Tail Moment (µm) vs. %DNA in Tail", AxesLabel →
             {"Olive Tail Moment (µm)" , "%DNA in Tail"}], "Output"]}]}],

     TextCell["Raw Data", "Section"],
     CellGroup[{
```

```
      Multicolumn[ExpressionCell[Dataset[countlist[[#]]]] & /@
        Range[Length[countlist]], 3]}]

  }];

reportsave[countlist_, focifilename_] :=
 Export[SystemDialogInput["FileSave",
   {TemplateApply[StringTemplate["`1`_`2`_+`3`_`4`_Comet_Report"],
     StringSplit[focifilename, {"_", "+"}][[{1, 2, 4, 5}]]],
    {"Tabular Formats (*.cvs, *.tsv)" → {"*.csv", "*.tsv"}}}],
  TableForm[Values[countlist][[All, {1, 3, 4, 5}]], TableHeadings → None]];

heatmap[img_, bin_] :=
 Module[{p, value, ass},
  p = ImageValuePositions[ColorNegate[Binarize[img, bin]], 0];
  value = Map[ImageValue[img, #] &, p];
  ass = AssociationThread[p, value];
  Graphics[{Point[Keys@ass, VertexColors → (Hue /@ Values@ass)]},
   ImageSize → {350, 350}, Axes → True]];

backgroundheatmap[img_, bin_, coords_, r_, xlen_, ylen_] :=
 Module[{p, value, ass, selectedkeys},
  p = ImageValuePositions[ColorNegate[Binarize[img, bin]], 0];
  value = Map[ImageValue[img, #] &, p];
  ass = AssociationThread[p, value];
  selectedkeys = KeySelect[ass, RegionMember[Rectangle[{coords[[1]] - r - xlen,
       coords[[2]] + r}, {coords[[1]] + r, coords[[2]] + r + ylen}]]];
 ];
backgroundsmooth[img_, bin_, coords_, r_, xlen_, sxlen_] :=
 Module[{p, value, ass, upperselectedkeys, lowerselectedkeys, uppersmoothed,
   lowersmoothed, smoothedass, intass, uppersmoothedass, lowersmoothedass},
  p = ImageValuePositions[ColorNegate[Binarize[img, bin]], 0];
  value = Map[ImageValue[img, #] &, p];
  ass = AssociationThread[p, value];
  upperselectedkeys = KeySelect[ass,
    RegionMember[Rectangle[{coords[[1]] - r - xlen, coords[[2]] + r + sxlen},
      {coords[[1]] + r, coords[[2]] + r * 2 + sxlen}]]];
  lowerselectedkeys = KeySelect[ass, RegionMember[
     Rectangle[{coords[[1]] - r - xlen, coords[[2]] - r * 2 - sxlen},
      {coords[[1]] + r, coords[[2]] - r - sxlen}]]];
  uppersmoothed = ListConvolve[{2, 2} / 6, Values[upperselectedkeys], {1, 1}];
  lowersmoothed = ListConvolve[{2, 2} / 6, Values[lowerselectedkeys], {1, 1}];
  uppersmoothedass =
```

```
   AssociationThread[Keys[upperselectedkeys], uppersmoothed];
  lowersmoothedass = AssociationThread[Keys[lowerselectedkeys],
    lowersmoothed];
  smoothedass = Append[uppersmoothedass, lowersmoothedass]
 ];

backgroundsmoothupper[img_, bin_, coords_, r_, xlen_, sxlen_] :=
 Module[{p, value, ass, upperselectedkeys, lowerselectedkeys, uppersmoothed,
   lowersmoothed, smoothedass, intass, uppersmoothedass, lowersmoothedass},
  p = ImageValuePositions[ColorNegate[Binarize[img, bin]], 0];
  value = Map[ImageValue[img, #] &, p];
  ass = AssociationThread[p, value];
  upperselectedkeys = KeySelect[ass,
    RegionMember[Rectangle[{coords[[1]] - r - xlen, coords[[2]] + r + sxlen},
      {coords[[1]] + r, coords[[2]] + r * 2 + sxlen}]]];
  uppersmoothed = ListConvolve[{2, 2} / 6, Values[upperselectedkeys], {1, 1}];
  uppersmoothedass =
   AssociationThread[Keys[upperselectedkeys], uppersmoothed];
 ];

backgroundsmoothlower[img_, bin_, coords_, r_, xlen_, sxlen_] :=
 Module[{p, value, ass, upperselectedkeys, lowerselectedkeys, uppersmoothed,
   lowersmoothed, smoothedass, intass, uppersmoothedass, lowersmoothedass},
  p = ImageValuePositions[ColorNegate[Binarize[img, bin]], 0];
  value = Map[ImageValue[img, #] &, p];
  ass = AssociationThread[p, value];
  lowerselectedkeys =
   KeySelect[ass, RegionMember[Rectangle[{coords[[1]] - r - xlen,
       coords[[2]] - r * 2 - sxlen}, {coords[[1]] + r, coords[[2]] - r - sxlen}]]];
  lowersmoothed = ListConvolve[{2, 2} / 6, Values[lowerselectedkeys], {1, 1}];
  lowersmoothedass =
   AssociationThread[Keys[lowerselectedkeys], lowersmoothed]
 ];



headplotdata[img_, bin_, coords_, r_, back_] :=
 Module[{hk, hv, ha, hsk, clb, pa},
  hk = ImageValuePositions[ColorNegate[Binarize[img, 0]], 0];
  hv = Map[ImageValue[img, #] &, hk];
  ha = AssociationThread[hk, hv];
  hsk = KeySelect[ha, RegionMember[Disk[coords, r]]];
  clb = getcolumnlength[back];
```

```
    pa = {#[[1]], integrateinetsity[#[[1]], hsk, 1] -
          ((integrateinetsity[#[[1]], back, 1]) / getcolenoverlap[#, clb, 1][[
              2]]) * #[[2]]} & /@ getcolumnlength[hsk]];

  tailplotdata[img_, bin_, coords_, r_, xlen_, sxlen_, sylen_, back_] :=
   Module[{tk, tv, ta, tsk, clb, pa},
    tk = ImageValuePositions[ColorNegate[Binarize[img, bin]], 0];
    tv = Map[ImageValue[img, #] &, tk];
    ta = AssociationThread[tk, tv];
    tsk = KeySelect[ta, RegionMember[Rectangle[{coords[[1]] - r - xlen,
          coords[[2]] - r - sylen}, {coords[[1]] - r, coords[[2]] + r + sxlen}]]];
    clb = getcolumnlength[back];
    pa = {#[[1]], integrateinetsity[#[[1]], tsk, 1] -
          ((integrateinetsity[#[[1]], back, 1]) / getcolenoverlap[#, clb, 1][[
              2]]) * #[[2]]} & /@ getcolumnlength[tsk]];


  getcolumns[x_] :=
   Flatten[DeleteDuplicates[#] & /@
     Select[Gather[Sort[Keys[x]][[All, 1]]], Length[#] > 1 &]];

  getcolumnlength[x_] :=
   Flatten[{DeleteDuplicates[#], Length[#]}] & /@
    Select[Gather[Sort[Keys[x]][[All, 1]]], Length[#] > 1 &];

  getcolenoverlap[in_, colen_, pos_] :=
   Flatten[Select[colen, #[[pos]] === in[[pos]] &]];


  integrateinetsity[y_, ass_, loc_] :=
   Integrate[Interpolation[Transpose[
       {Range[Length[Values[KeySort[KeySelect[ass, #[[loc]] === y &]]]]],
        Values[KeySelect[ass, #[[loc]] === y &]]}], InterpolationOrder → 2][w],
    {w, 1, Length[Values[KeySort[KeySelect[ass, #[[loc]] === y &]]]]}];

)]]]
```

Introduction and results

We previously generated Dot1L knockout (*Dot1L*$^{-/-}$) mice to study the function of Dot1L during embryogenesis. A majority of *Dot1L*$^{-/-}$ mice die at embryonic day 10.5 (E10.5) with defects in erythroid progenitor development, G1/G0 arrest, and increased apoptosis when treated with erythroid growth factors [162]. The anemic phenotype in *Dot1L* $^{-/-}$ mice is hypothesized to be a result of the loss of an erythroid specific transcriptional program [106, 107, 162, 164-167]. However, an accumulating body of work indicates that Dot1L plays a role in genome integrity and DNA repair[112-116, 126, 127]. Thus, we propose an alternative hypothesis: the loss of erythroid lineage cells in *Dot1L*$^{-/-}$ mice is a function of its role in DNA repair. Hematopoietic stem cells (HSC) accumulate DNA damage overtime through the generation of reactive oxygen species (ROS) [168, 169]. The *in vivo* cytokine-induced mobilization of quiescent HSCs forces their entry into the cell cycle and forces a switch from the NHEJ to the HR repair pathway [170]. Our previous work demonstrated that E10.5 *Dot1L*$^{-/-}$ mice have similar numbers of hematopoietic progenitors as wild-type littermates, however, cytokine-induced differentiation promoted apoptosis [162]. Several reports indicate that compromised DNA repair pathways decrease the ability of HSC to self-renew and differentiate [171-174]. To interrogate whether hematopoietic progenitors derived from *Dot1L*$^{-/-}$ mice accumulate DNA damage, cells from E10.5 yolk sacs were cultured in media containing the cytokines SCF, IL-3, IL-6, and EPO, which promote the growth of erythroid and myeloid progenitors. After 4 days of growth, we harvested the cells and analyzed endogenous DNA damage. Utilizing an alkaline comet assay, we observed increased DNA DSBs and single-strand breaks (SSB) in *Dot1L*$^{-/-}$ hematopoietic progenitors when compared to wild-type littermates (Figure 5.1). The observation that *Dot1L*$^{-/-}$ embryos are anemic is the most noticeable phenotype; however, upon closer inspection the extra-

embryonic vasculature of the yolk sac contains drastic alterations [162, 175]. Disrupted architecture of the HSC niche in the yolk sac of *Dot1L*$^{-/-}$ embryos could contribute to their anemic phenotype [176]. Therefore, we tested if alternations in proliferation and DNA damage were limited only to the hematopoietic linage.

Our previous work demonstrated that cells isolated from *Dot1L*$^{-/-}$ yolk sacs have a G1/G0 arrest when stimulated with growth cytokines [162]. We tested if cells isolated from the embryo proper of E10.5 *Dot1L*$^{-/-}$ mice had similar cell cycle delay using a microscopy cell cycle analysis method (MANA, Chapter IV) [155]. We compared the overall proportion of *Dot1L*$^{-/-}$ and wild-type cells in each phase of the cell cycle and observed more *Dot1L*$^{-/-}$ cells in the G1/G0 phase and fewer cells in the G2/M phase (Figure 5.2a). The G1/G0 accumulation is similar to our previous results for *Dot1L* $^{-/-}$ erythroid differentiation culture [162]. We also observed that *Dot1L*$^{-/-}$ cells proliferate at slower rates than wild-type cells, and stop dividing after three passages (Figure 5.2b). We corroborated the proliferation defects using a microscopy based BrdU incorporation assay, in which we pulsed cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos for 24 hours . Cells from *Dot1L*$^{-/-}$ embryos showed less BrdU positivity when compared to wild-type littermates four days after isolation (Figure 5.3). Based on these results, we generated two hypothesis: first, there exists a bona fide G1/G0 arrest in *Dot1L*$^{-/-}$ embryonic cells thereby limiting the pool of cells that are able enter S/G2/M phases; second, a subset of *Dot1L*$^{-/-}$ cells fail to exit S/G2/M properly and never renter the cell cycle. If cell loss occurs in S/G2/M phase, it would appear as though a G1/G0 arrest was occurring as the proportion of cells remaining inflates the G1/G0 peak.

Implementation of MANA relies on training a deep learning neural network for nuclei identification (see Chapter IV). Human curated nuclei images were used as training data for the

network (see Chapter IV, Figure 4.5). We examined the output of the neural network used for MANA and found that abnormally shaped nuclei were excluded from the cell cycle analysis. Characterization and quantification of these nuclei reveled that cells isolated from E10.5 *Dot1L-/-* embryos have increased abnormal nuclear boundaries with blebs, chromosomal bridges, and mitotic catastrophes (Figure 5.4a) [177]. This data supports the hypothesis that the decrease in G2/M phase population is a result of incomplete G2/M exit. During our assessment of the MANA neural network output we also observed that nuclei from *Dot1L*$^{-/-}$ embryos contain increases in micronuclei abundance, a marker of genomic instability (Figure 5.4b)[178, 179]. Replicative stress is hypothesized to be a source for micronuclei formation, and bypassing the intra-S-phase check point can lead to the accumulation of chromosomal breaks and abnormalities [178, 180]. It has been reported that the yeast mutant strain, *ΔDot1,* has a defective S-phase check point [113]. It is conceivable that cells from E10.5 *Dot1L*$^{-/-}$ embryos accumulate DNA damage during the S-phase that goes unrepaired resulting in a highly unstable genome and a loss of cells during G2/M.

We therefore tested if cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos fail to repair DNA following genotoxic stress. We used low dose ionizing radiation (IR) to introduce DNA damage, and assayed for DNA repair using an alkaline comet assay. Cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos and wild-type littermates have similar amounts of DSBs and SSB 24 hours post IR (Figure 5.5). However, 72 and 96 hours post IR cells from E10.5 *Dot1L*$^{-/-}$ embryos accumulate more DNA damage than wild-type cells (Figure 5.5). This indicates that after exposure to DNA damage E10.5 *Dot1L*$^{-/-}$ cells fail to completely repair DNA breaks. The persistence of unrepaired DNA damage can lead to genomic instability, cell cycle arrest, and apoptosis. To that end we tested if the presence of unrepaired DNA damage could lead to growth arrest. Using a

microscopy based BrdU incorporation assay we pulsed cells isolated from E10.5 embryos for 24 hours and observed that 96 hours after IR treatment both *Dot1L* $^{-/-}$ and wild-type cells have decreased proliferation rates as compared to non-IR treated cells (Figure 5.6, Figure 5.3). Interestingly even though cells from E10.5 *Dot1L*$^{-/-}$ embryos have more DNA damage at 96 hours post IR, proliferation persists, albeit at a slower rate (Figure 5.6). This observation suggests that cells from E10.5 *Dot1L*$^{-/-}$ still proliferate in the presence of large amounts of DNA damage, which we hypothesize is a mechanism for why they have increased endogenous DNA damage, as we observed in erythroid lineage cells isolated from *Dot1L*$^{-/-}$ mice (Figure 5.1).

An accumulation of unrepaired DNA indicates that cells from *Dot1L*$^{-/-}$ embryos have defects in DNA damage repair (DDR) pathways. It has been hypothesized that Dot1L might have a roll in the recruitment of NHEJ repair factor 53BP1, however a mechanism for direct recruitment of DNA repair factors remains to be identified [115, 116, 126, 127]. Dot1L has primarily been studied for its role as a transcriptional activator and it is possible that the accumulation of DNA damage in cells isolated from *Dot1L-/-* embryos is a result of decreased expression of DDR genes [108]. We performed a comprehensive analysis of DDR pathway factors using cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos and wild-type littermates. Cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos have elevated expression of several factors involved in sensing DNA damage (Figure 5.7a).

Chd1l is a factor known to interact wtih poly(ADP-ribose) (PAR), histones, and facilitate nucleosome remolding and DDR [181, 182]. Parp1 (PAR polymerase 1) is responsible for deposition of PAR at break sites and Parg is a glycohydrolase responsible for the removal of PAR preventing accumulation during prolonged replicative stress, both are necessary for DDR to take place [183, 184]. The histone variant H2afz (H2AZ) is exchanged in nucleosomes

surrounding the DSB sites and restricts the end resection machinery [136]. The upregulation of DNA damage sensors in cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos suggests that transcripts necessary for sensing DNA damage are present. At the protein level, E10.5 *Dot1L*$^{-/-}$ embryonic cells form DDR foci for the DNA damage sensor γH2A.X after treatment with IR, indicating they are capable of sensing DSBs (Figure 5.8).

We also observed that cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos have increased expression of several transcripts necessary for HR repair (Figure 5.7b). Of note is the increase in expression of Brca2, Exo1, Rbbp8, Rmi1, Top3a, Blm, Slx4, Slx1b, and Gen1. Brca2 interacts with ssDNA and facilitates the loading of Rad51 onto ssDNA following DNA damage [185]. Exo1 is an endonuclease that responsible for long-range end resection events following DNA damage [186]. Rbbp8, (aka CtIP), is an endonuclease that is responsible for short-range end resection events following DNA damage [88]. Blm, Rmi1, and Top3a form a complex that enhance end resection through DNA helicase activity [187, 188]. Slx4, Slx1b, and Gen1 are involved in holiday junction resolution following DNA damage [189-191].

We examined the ability of cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos to form DDR foci for the HR factor Rad51 and we observed a delay in its accumulation following IR treatment compared to wild type littermates (Figure 5.9). The expression of Rad51 is reduced in wild-type cells isolated from E10.5 embryos after exposure to IR; however, this does not preclude the formation of DDR foci (Figure 5.7b, Figure 5.9). The expression of Rad51 in cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos is at a similar level to wild-type littermates after exposure to IR (Figure 5.7b). The reduction in transcript level for all genes analyzed following IR exposure (Figure 5.7) is not without precedent. It has been reported that RNA polymerase II is inhibited following DNA damage, leading to a reduction in transcript levels [192]. We also examined the expression

of other genes involved in the NHEJ and alternative NHEJ (alt-NHEJ) pathway (Figure 5.7c). Similar to the genes in the DNA damage sensing and HR repair pathways cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos have elevated expression of NHEJ and alt-NHEJ pathway transcripts (Figure 5.7c).

We observed the elevated expression of Lig4, Trp53bp1, Xrcc4, Lig3, and Rad52 (Figure 5.7c). Classically considered a promoter of NHEJ, Trp53bp1 is thought to prevent Brca1/CtIP from loading onto breaks, thereby promoting NHEJ though Ku70/Ku80 binding [148]. However, recent work points to a role for Trp53bp1 in HR repair, whereby it prevents hyper end resection and single strand annealing (SSA) repair via Rad52 by promoting Rad51 assembly on ssDNA [77]. It is an intriguing correlation that Rad52 expression is elevated along with Exo1, another factor involved in hyper end resection and SSA repair (Fig 2d). The yeast mutant strain, *Δdot1,* has a hyper end resection phenotype [125]. We analyzed Trp53bp1 foci formation in cells isolated from E10.5 *Dot1L*$^{-/-}$ cells that were treated with IR (Figure 5.10). Surprisingly, and counter to previous reports, we did not observe a loss of Trp53bp1 foci following IR treatment 2 hours post IR, indicating that some aspect of NHEJ repair pathway is intact. Our data indicate that cells from *Dot1L*$^{-/-}$ embryo can sense and respond to DNA damage, however, the response is altered with a delay in the formation of Rad51 damage following IR treatment (Figure 5.9). We hypothesize that the alteration in DDR repair factor recruitment can shift the balance of DDR pathway utilization.

Dot1L is the only reported H3K79 histone methyltransferase. The methyltransferase activity of Dot1L and regions of its C-terminus have been hypothesized to be involved in DNA repair [116, 126]. In our *Dot1L*$^{-/-}$ mouse model, expression from the *Dot1L* locus is ablated using a gene trap, thereby preventing us from separating Dot1L enzymatic activity from protein-

protein interaction activity [162]. We sought to isolate the enzymatic function of Dot1L in DNA repair from other protein functions using a genetic approach. We utilized the CRISPR/Cas9 system to generate two DOT1L mutant HEK293T cell lines; $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ (Figure 5.12, Figure 5.13) [193, 194]. We chose to target Y312A because it is a critical residue for methylation, and satisfied the CRISPR/Cas9 targeting constraints [195, 196]. We designed an eGFP-P2A-Cas9 construct to allow for transient expression and cell selection via FACS to avoid stable integration, and expression of the Cas9 construct (Figure 5.11a). We utilized a 150 nucleotide single stranded oligonucleotide for homology directed repair to introduce nucleotide substitutions into the *DOT1L* locus that converted the nucleic acid sequence corresponding to amino acid Y312 to either three tandem stop codons or an alanine while simultaneously generating a ubiquitous restriction site DdeI used for screening (Figure 5.11b, Figure 5.12, Figure 5.13). Both $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ results in loss of H3K79 di- and tri-methylation marks, while only $DOT1L^{Stop}$ reduces DOT1L protein level (Figure 5.11b). We were surprised that HEK29T cells remained viable after targeting, as our work with primary mouse cells and other immortalized human cell lines (data not shown) led us to anticipate that reduction of DOT1L protein or inhibition of enzymatic activity would result in cell death.

In DSB break repair two factors influence repair pathway utilization: the stage of the cell cycle in which the DSB occurs, and how quickly DNA damage repair factors assemble at the break site [71, 197, 198]. An initial step of DNA double strand break repair factor assembly begins with the MRN (Mre11-Rad50-Nbs1) complex sensing the break site, and phosphorylation and activation of ATM [199]. Activated ATM in turn phosphorylates the histone variant H2A.X at the site of the break and in concert with H2AZ, sets the boundary for the DNA damage repair machinery [136, 200]. The phosphorylation of H2A.X is required for the recruitment of MDC1

to site of the break [67]. MDC1 binding then leads to clearance of proteins surrounding the break site and, depending on the timing, either HR or NHEJ repair will be used [71]. For example, 53BP1 can promote NHEJ at DNA breaks by out-competing the recruitment of HR repair factors such as BRCA1 [57, 71]. To gain insight into the mechanism of DOT1L effect on DDR factor loading, we studied the kinetics of 53BP1 and BRCA1 DNA damage foci formation (Figure 5.14). We treated wild-type, $DOT1L^{Stop}$, and $DOT1L^{Y312A}$ cells with IR and analyzed 53BP1 foci formation at different time points (Figure 5.14). We observed that $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ cells have reduced 53BP1 foci compared to wild-type HEK293T cells starting at 30 minutes for $DOT1L^{Y312A}$ and 60 minutes for $DOT1L^{Stop}$ (Figure 5.14a). BRCA1 foci levels for $DOT1L^{Y312A}$ are significantly less at 60 minutes post IR while little effect on BRCA1 foci is observed in $DOT1L^{Stop}$ mutants (Figure 5.14b). We also observed that in $DOT1L^{Y312A}$ mutants exposed to IR, BRCA1 is predominately localized to the cytoplasm, this could account for its reduction in DNA damage foci (data not shown). 53BP1 and BRCA1 are classically viewed as having a competing roll in DNA damage repair and work to limit each other at break sites [201]. However, recent studies have added evidence for 53BP1 working to facilitate HR by limiting end resection {Ochs, 2016 #237. We compared the co-localization of 53BP1 and BRCA1 following IR treatment. We observed a decrease 53BP1/BRCA1 co-localization events in $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutant cell lines at later times post IR. (Figure 5.14c). The alteration to 53BP1 and BRCA1foci formation and localization in *DOT1L* mutants could lead to alterations in repair pathway utilization.

To gain functional insight into the observed foci kinetic differences in wild-type cells, $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ HEK293T cells we utilized three different systems to monitor DDR pathway utilization: DR-GFP for HR, E5J-GFP, and E2J-GFP for NHEJ and alt-NHEJ

{Bennardo, 2008 #238;Pierce, 1999 #113}. We observed that 48 hours after introducing a break into the DR-GFP cassette, both *DOT1L*^{Stop} and *DOT1L*^{Y312A} have reduced levels of GFP positive cells indicating a loss of HR repair activity (Figure 5.15a). This result correlates with the delay in BRCA1 foci formation we observed in *DOT1L*^{Stop} and *DOT1L*^{Y312A} (Figure 5.14b). Next we examined NHEJ repair activity using wild-type, *DOT1L*^{Stop}, and *DOT1L*^{Y312A} cells lines stability expressing the E5J-GFP plasmid. We observed the *DOT1L*^{Stop} mutant could still process breaks using NHEJ, but interestingly there was a decrease in NHEJ repair in the *DOT1L*^{Y312A} mutant (Figure 5.15b). This suggests that at some level, NHEJ repair requires a DOT1L protein that is enzymatically active or it directly requires H3K79 methylation. Our DDR foci show that both in *DOT1L*^{Stop} and *DOT1L*^{Y312A} form 53BP1 foci similar to wild-type, but have altered BRCA1 foci formation kinetics. We hypothesize that the attenuation of BRCA1 foci formation and the loss of co-localization in *DOT1L*^{Stop} and *DOT1L*^{Y312A} leads to a favoring of NHEJ repair pathway. We also examined alt-NHEJ repair activity using wild-type, *DOT1L*^{Stop}, and *DOT1L*^{Y312A} cells lines stability expressing the E2J-GFP plasmid, which measures alt-NHEJ activity. We observed *DOT1L*^{Stop} had increased amounts alt-NHEJ activity compared to wild-type (Figure 5.15c). The increase in alt-NHEJ is consistent with our observations of upregulation of in NHEJ pathway components and end resection machinery in cells isolated from E10.5 *Dot1L*$^{-/-}$ embryos (Figure 5.7b,c). Our data point to a mechanism whereby in the absence of DOT1L, cells utilize error prone DNA repair pathways.

We were interested in understanding how DOT1L and DOT1L mediated H3K79 methylation activity might influence histone dynamics at DSB sites. DOT1L mediated H3K79me is a unique epigenetic mark as it occurs on the core of histone. It is hypothesized that H3K79me can influence the how tightly the histone octamer can associate with DNA because H3K79me

leads to open, relaxed chromatin and in its absence, the DNA is more tightly associated with histones. We wanted to test this hypothesis in the context of DNA damage-induced histone dynamics using a nucleosome stability assay. In this assay, we analyze histones that remain associated with a normalized about DNA following IR in the presence of high salt (Figure 5.16). Recent work has indicated that histone exchange of H2AZ at sites of DSB can influence DNA repair pathway choice [136]. Therefore, we tested the stability of H2AZ in the genome of wild-type, and $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ HEK293T cells (Figure 5.16a). We observed that 60 minutes post IR, H2AZ is easily removed from DNA of wild-type HEK293T cells and is more stable in $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutants. (Figure 5.16a). The stability of H2AZ in wild-type chromatin increases with time, and by 240 minutes post IR H2AZ reaches a plateau of stability (Figure 5.16a). The stability of H2AZ in $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutants does not change at between –IR and up to 240 minutes post IR treatment (Figure 5.16a). This leads us to hypothesize that in $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutants H2AZ is more stably associated with the DNA and is not easily removed following IR treatment.

An epigenetic mark that has been associated with relaxed and open chromatin following DNA damage is H4 acetylation [144]. It is hypothesized that the exchange of H2AZ at the site of DNA breaks leads to a packed chromatin state and the subsequent removal of leads to H4 acetylation, transitioning the chromatin into a more relaxed state permissive for HR repair [144]. Using a nucleosome stability assay, we measured the stability of H4 acetylated histones following IR treatment (Figure 5.16b). In wild-type, $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutant HEK293T cells, we observe a decrease in H4 acetylated histones following IR (Figure 5.16b). This result can be two ways; first, H4 acetylated histones are associated with nucleosomes less stabl after IR, or second H4 is rapidly deacetylated following IR. We also analyzed H3 stability

following IR treatment and observed that in wild-type HEK293T cells, H3 is less stable 60 minutes following IR treatment (Figure 5.16c). This suggests a general phenomenon that genome-wide stability of histones is altered in Dot1L deficient cells, perhaps allowing the histone exchange necessary for DSB repair to take place. It is of note that in *DOT1L*^{Stop} and *DOT1L*^{Y312A} mutants, H3 is more stable than in the wild-type cell line (Figure 5.16c). This observation suggests that in cells lacking a function DOT1L have histones that are more tightly pack and not easily removed. The histone dynamics measured by stability assays are only capable of global detection, and information about local histone dynamics around the break site are lost. To address this limitation we developed an assay that can monitor the histone dynamics at break sites: BeL-ChIP (broken end ligation ChIP) (Figure 17a, Chapter II for details). In this assay we ligated a hairpin adapter to broken ends of DNA following IR treatment, followed by a ChIP procedure, and ligated a second adapter following shearing of the genomic DNA. Only fragments that contain both adaptors were amplified, thereby allowing for identification histones present at the break site. We performed the BeL-ChIP on H2AZ 60 minutes post IR in wild-type, *DOT1L*^{Stop} and *DOT1L*^{Y312A} mutant HEK293T cells. We found that in *DOT1L*^{Stop} and *DOT1L*^{Y312A} mutants, H2AZ is depleted at the broken ends of DNA (Figure 5.17b). Our interpretation of this result is that histone remolding is defective in *DOT1L*^{Stop} and *DOT1L*^{Y312A} mutants and H2AZ cannot be efficiently removed and added back to the genome.

We also analyzed H4 acetylation using BeL-ChIP because H4 acetylation has been shown to correlate with H2AZ exchange. There were no significant differences in H4 acetylation at DSB sites suggesting that our results in Figure 5.16b might represent a global deacetylation event and histones surrounding the break site can still be acetylated. If histones around the break site can still be acetylated it, could mean that at some level 53BP1 binding could be inhibited at

the break site [201]. This makes sense for the process of DSB repair to take place, as it is hypothesized that if 53BP1 is able to flank the DSB site, end resection will be blocked [202]. Perhaps H4 acetylation allows for some end resection to take place, allowing for the removal of histones around the DSB site.

The consequence of H2AZ retention may be to restrict the end resection machinery [136]. In the absence of H2AZ, end resection is thought to be over-active and favor the generation of long ssDNA that is a poor substrate for HR and NHEJ [136]. In yeast, it has been observed that Dot1 mutants have a hyper-end resection phenotype, but this has not been reported in mammals [125]. We designed an assay to measure the length of ssDNA in a sequence independent manner (Chapter III for details). Briefly, we enriched for ssDNA following IR treatment, converted to dsDNA, ligated adaptors, and amplified the products. We observed an abundance of ssDNA fragments between 0.5kb to 3kb in $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ mutants (Figure 5.18). This indicates that in the absence of DOT1L, cells generate more ssDNA. $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ display different lengths of ssDNA generation with $DOT1L^{Stop}$ generating abundant, but smaller ssDNA fragments while $DOT1L^{Y312A}$ generates longer fragments. In both mutants, generation of ssDNA is a poor substrate for HR and could be why both $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ have defects in HR. The differing effects both $DOT1L^{Stop}$ and $DOT1L^{Y312A}$ have on NHEJ and alt-NHEJ could be explained in that $DOT1L^{Y312A}$ have longer ssDNA fragments overall that block NHEJ, but not alt-NHEJ. $DOT1L^{Stop}$ have shorter ssDNA fragments that can still be processed by NHEJ machinery, but are better substrates for alt-NHEJ. This points to a continuum of ssDNA generation and DNA repair pathway choice in DOT1L mutants.

Discussion

These results provide novel insights into the molecular function of Dot1L/DOT1L in DSB repair, and identify the importance of Dot1L in histone dynamics at the DSB sites. We identify a role for DOT1L and its methylation activity in promoting an open chromatin state that is permissive for H2AZ exchange, and allow HR DDR to take place. We propose the following model based on our observations and expand upon the model put forth by Brandon Price (Figure 19.). . Under normal conditions DOT1L facilitates exchange of H2AZ at sites of DNA damage by marking histones with H3K79me. This transitions chromatin into a relax conformation by reducing the affinity of DNA to the core histone. In the absence of DOT1L and H3K79me H2AZ cannot be exchanged at sites of DNA damage efficiently do the higher affinity of DNA to the nucleosome. The reduction of H2AZ at DNA breaks leads to the generation of long ssDNA favors DNA repair pathways that are highly mutagenic and increase genomic instability. Our results have major implications in the use of DOT1L inhibitors as promising therapeutic for leukemia, and also identify other DNA repair pathways that are utilized in the absence of a functioning DOT1L. We predict that using DOT1L inhibitors in combination with inhibitors of NHEJ could increase the efficacy of cancer therapeutics.

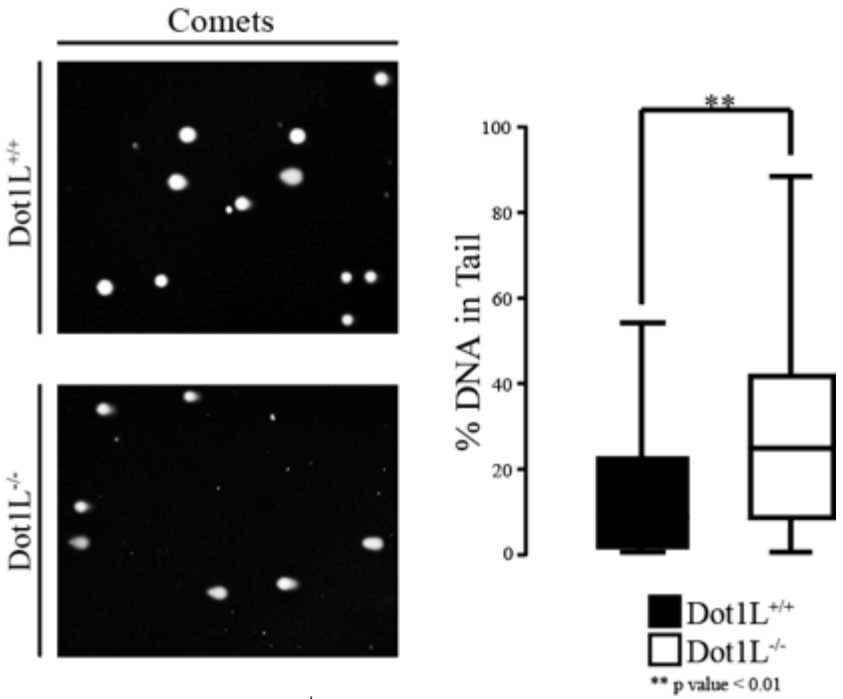


Figure 5.1: Cells isolated from Dot1L$^{-/-}$ yolks sacs accumulate DNA damage. Cells from Dot1L$^{+/+}$ and Dot1L$^{-/-}$ yolk sacs were cultured in erythroid differentiation media and then subjected to an alkaline comet assay. Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, p-value < 0.01, two tailed t-test.

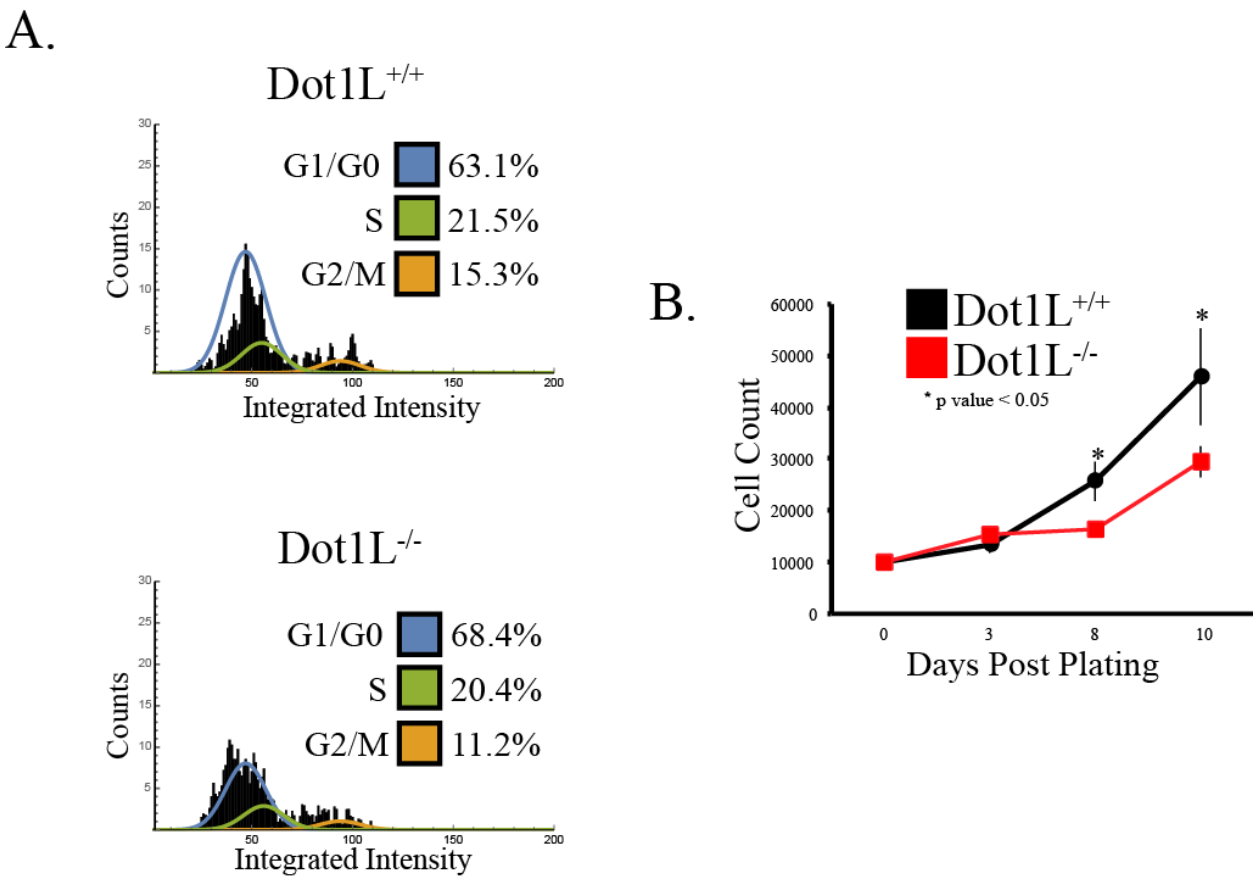


Figure 5.2: Cells isolated from Dot1L$^{-/-}$ embryos have altered cell cycle profiles and growth kinetics. A. The cycle profiles for Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells isolated from E10.5 embryos were calculated using MANA (See Chapter IV). B. The growth kinetics of cells isolated from Dot1L$^{+/+}$ and Dot1L$^{-/-}$ E10.5 embryos. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, p-value < 0.01, two tailed t-test)

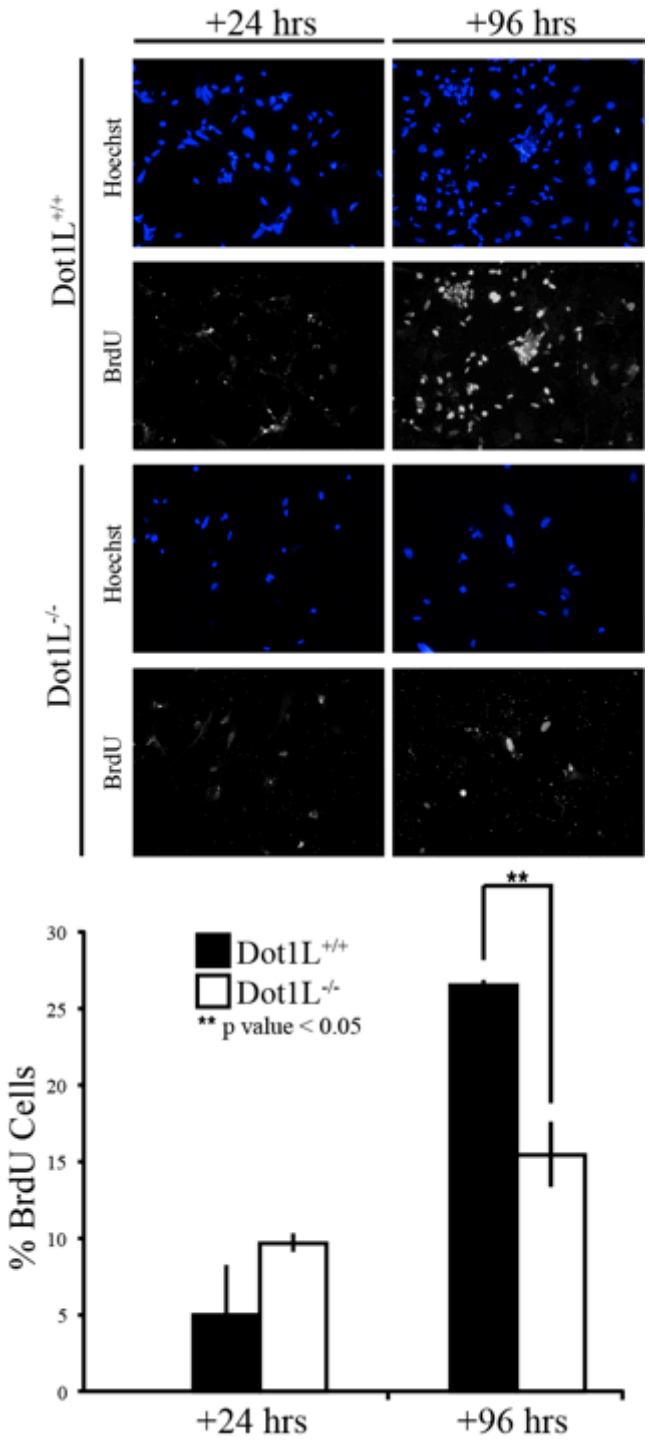


Figure 5.3: Cells isolated from Dot1L$^{-/-}$ embryos have fewer cells that complete S and G2/M phases. BrdU assay on cells isolated from Dot1L$^{+/+}$ and Dot1L$^{-/-}$ embryos. Cells were pulsed with BrdU for 24 hrs and BrdU incorporation was measured 24 and 96 hours after isolation from embryos. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, p-value < 0.05, two tailed t-test)

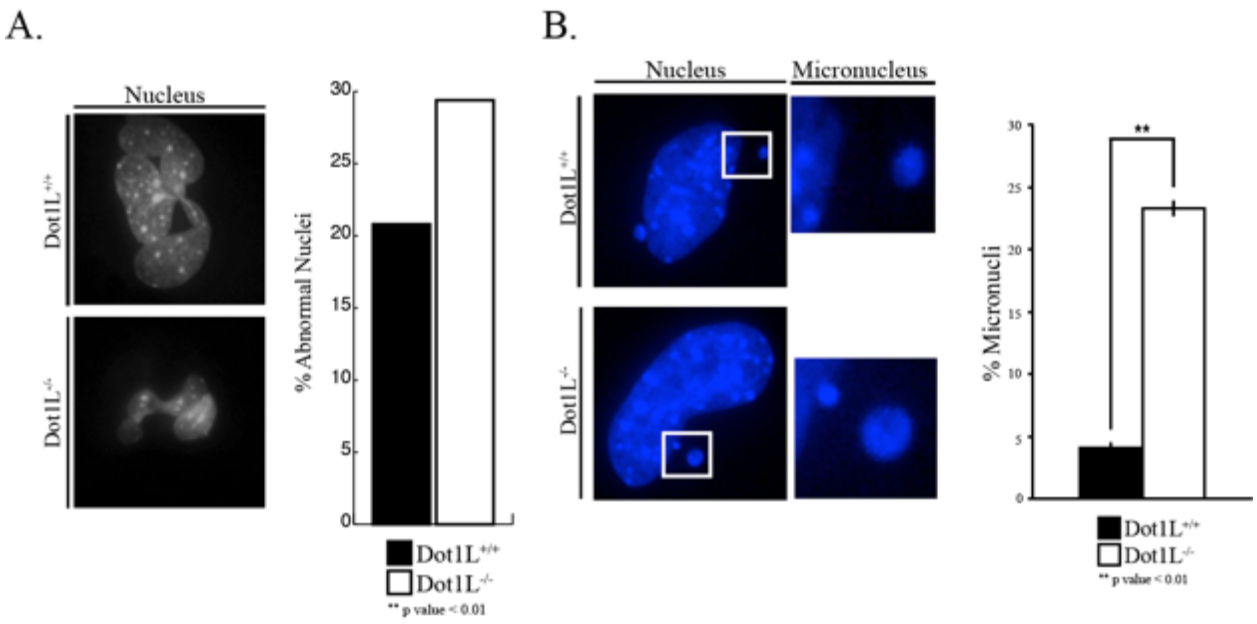


Figure 5.4: Cells isolated from Dot1L$^{-/-}$ embryos have high abnormal nuclei and micronuclei. A. Abnormally shaped nuclei were identified from the MANA output for both Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells isolated from E10.5 embryos. B. Micronuclei abundance was calculated for Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells isolated from E10.5 embryos. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, p-value < 0.01, two tailed t-test)

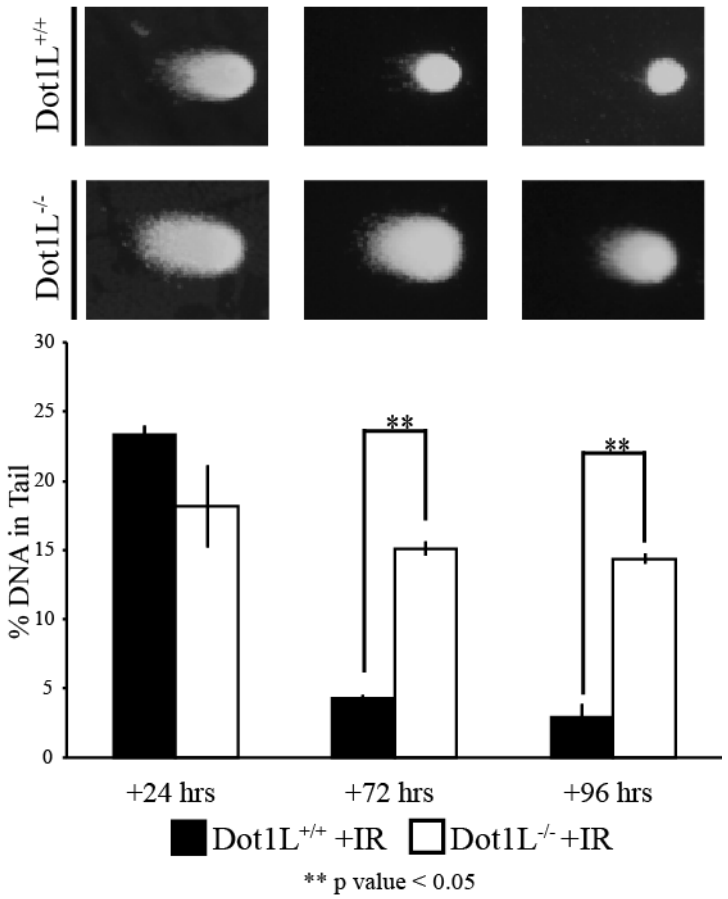


Figure 5.5: Cells isolated from Dot1L$^{-/-}$ E10.5 embryos fail to repair induced DNA damage. Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells were exposed to 2 Gy of ionizing radiation and measured for DNA damage levels using a alkaline comet assay 24, 72, and 96 hours post ionizing radiation treatment. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, p-value < 0.05, two tailed t-test)

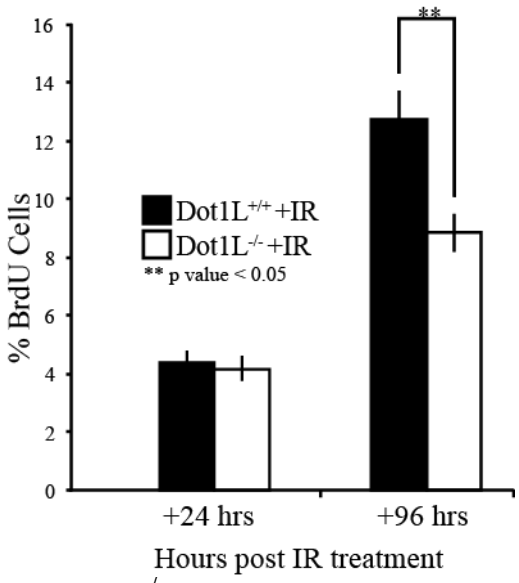


Figure 5.6: Cells isolated from Dot1L$^{-/-}$ E10.5 embryos continue to proliferate after ionizing radiation treatment. BrdU assay on cells isolated from Dot1L$^{+/+}$ and Dot1L$^{-/-}$ embryos exposed to 2 Gy of ionizing radiation. Cells were pulsed with BrdU for 24 hrs and BrdU incorporation was measured 24 and 96 hours after isolation from embryos. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, p-value < 0.05, two tailed t-test)

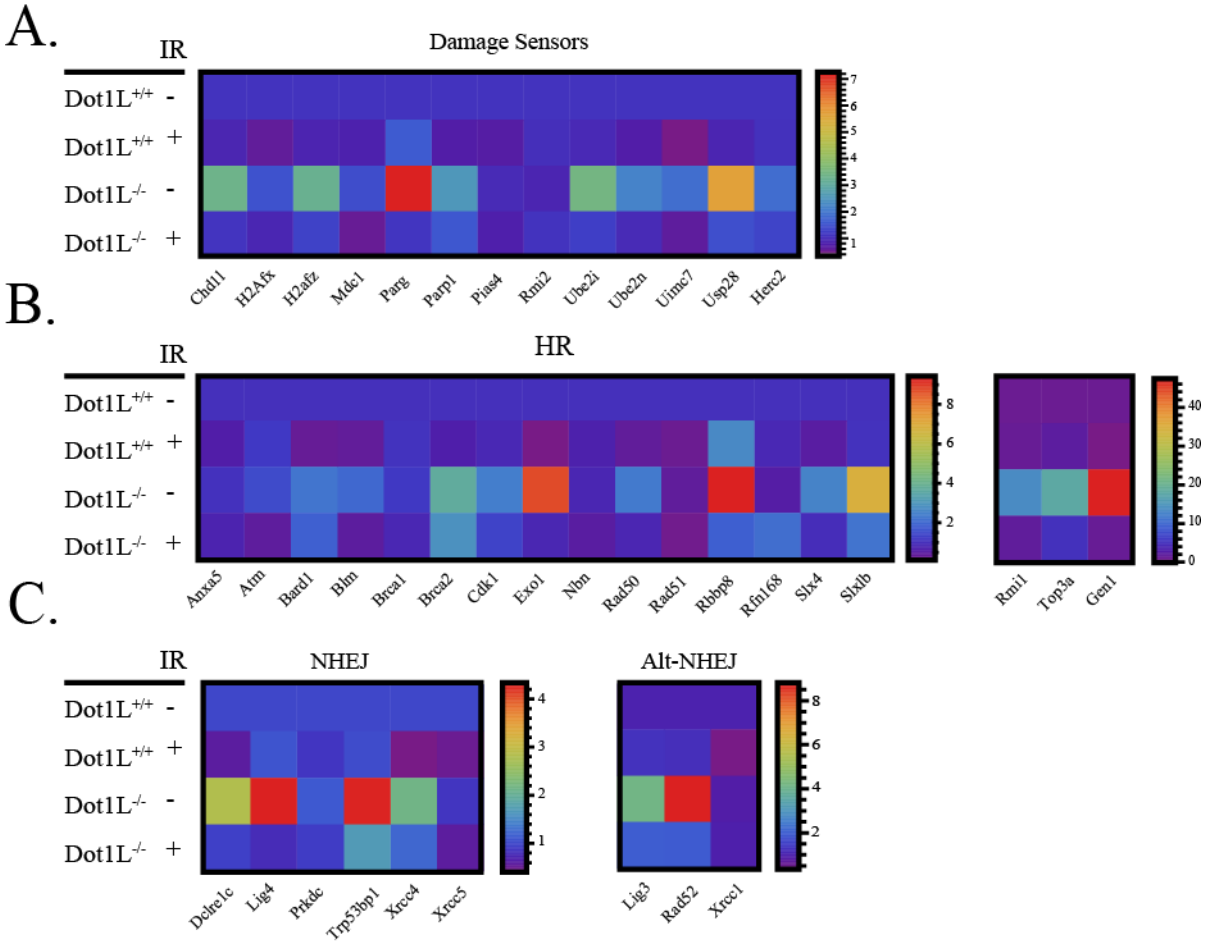


Figure 5.7: Cells isolated from Dot1L$^{-/-}$ E10.5 embryos have elevated expression of genes involved in DNA damage repair. qPCR was performed on RNA isolated from Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells extracted from E10.5 embryos. A. DNA damage sensor genes. B. Homologous recombination genes. C. Non homologous end joining and alter non homologous end joining genes. Fold change is colored coded, Blue is minimum, red is maximum. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3).

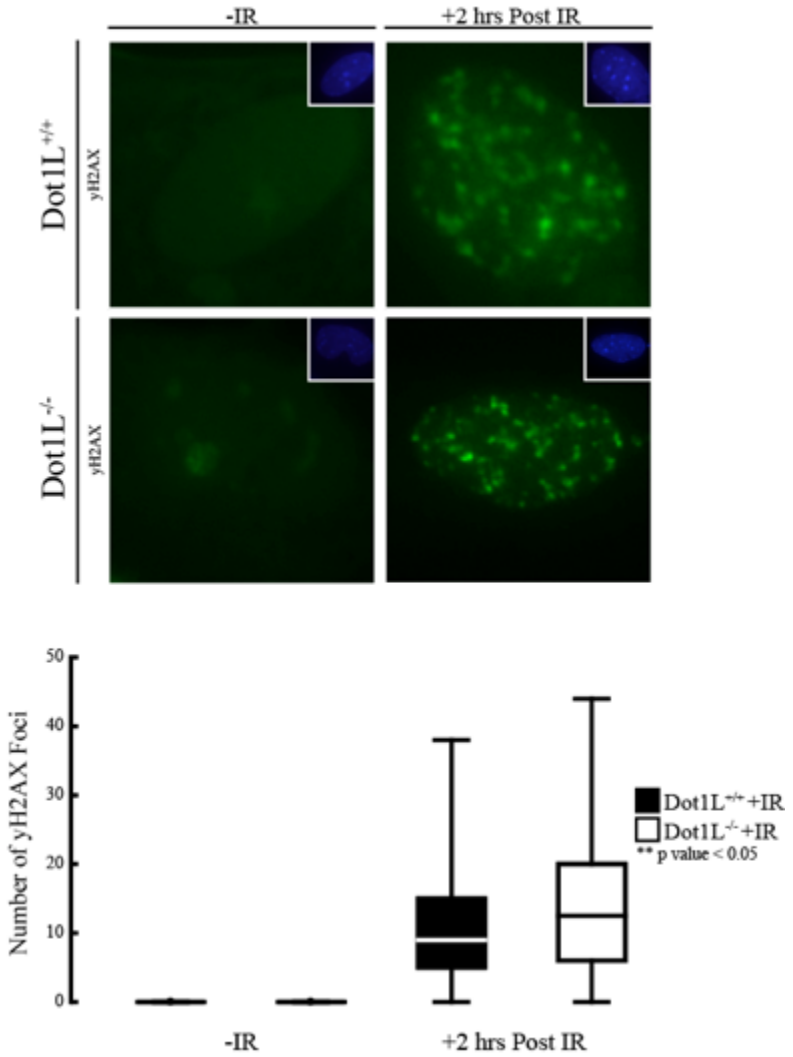


Figure 5.8: Cells isolated from Dot1L$^{-/-}$ E10.5 embryos have normal γH2AX foci formation. Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells were exposed to 10 Gy of ionizing radiation and γH2AX foci was quantified 2 hours later. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, >100 cells per n, p value < 0.05, two tailed t-test).

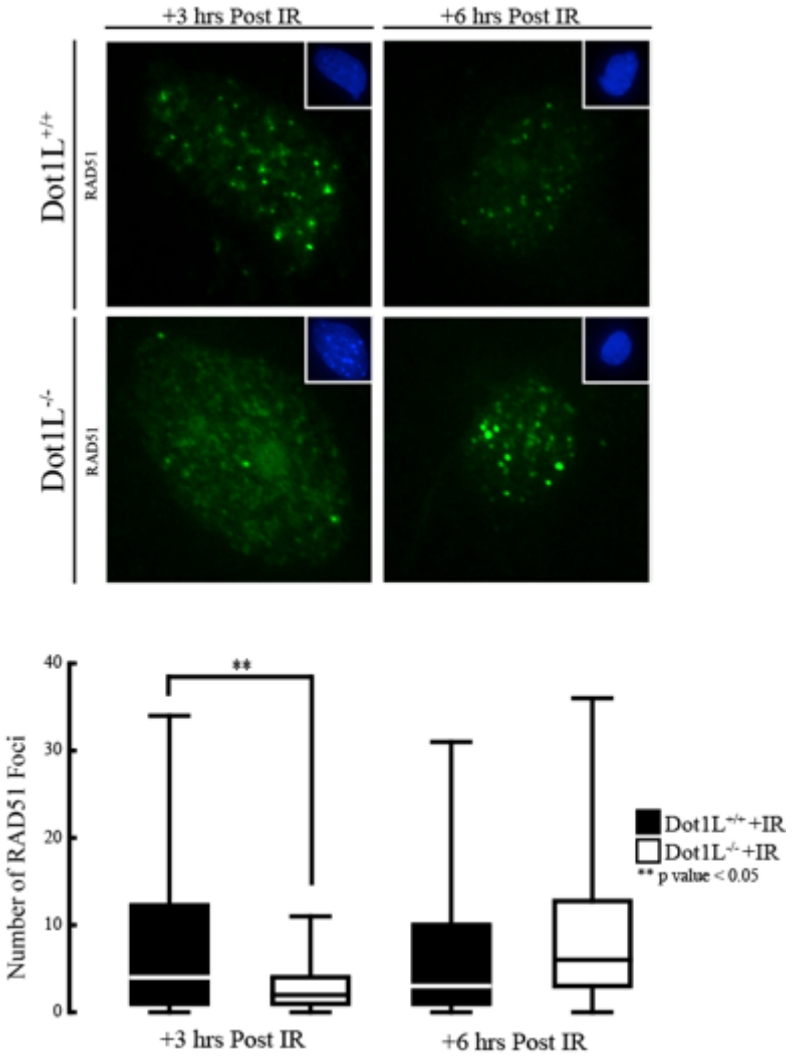


Figure 5.9: Cells isolated from Dot1L$^{-/-}$ E10.5 embryos have normal RAD51 foci formation. Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells were exposed to 10 Gy of ionizing radiation and RAD51 foci was quantified 3 and 6 hours later. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, >100 cells per n, p value < 0.05, two tailed t-test).

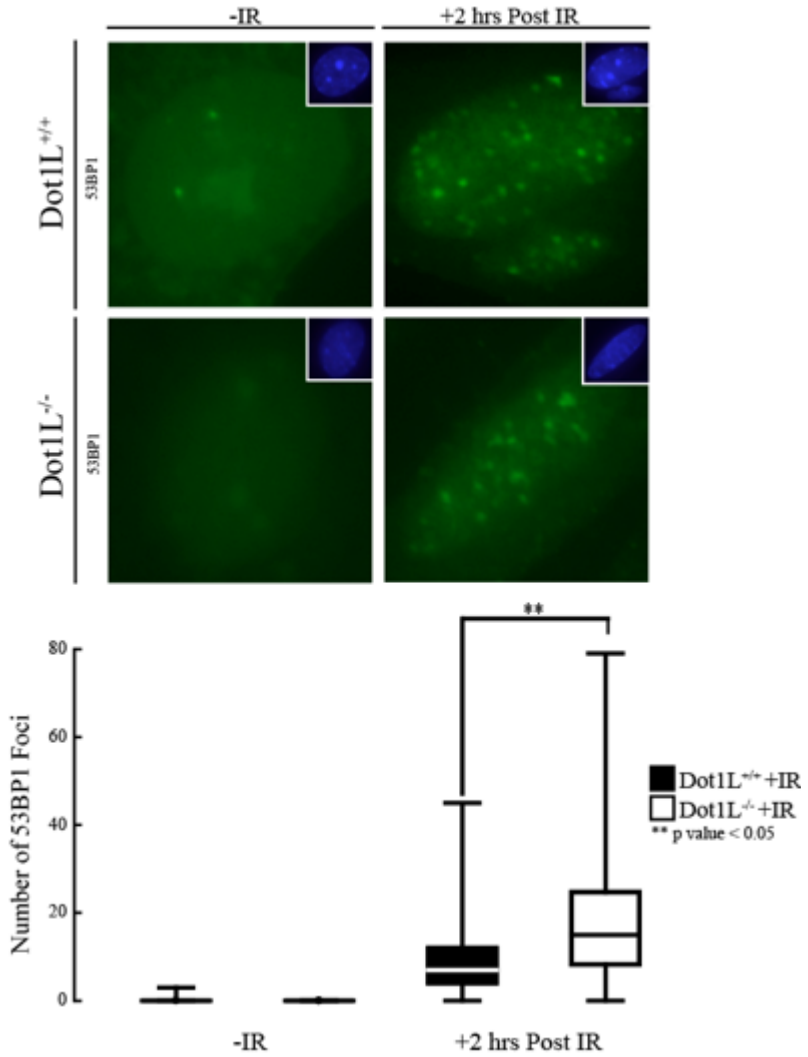


Figure 5.10: Cells isolated from Dot1L$^{-/-}$ E10.5 embryos have normal 53BP1 foci formation. Dot1L$^{+/+}$ and Dot1L$^{-/-}$ cells were exposed to 10 Gy of ionizing radiation and 53BP1 foci was quantified 2 hours later. (Dot1L$^{+/+}$ n=3, Dot1L$^{-/-}$ n=3, > 100 cells per n, p value < 0.05, two tailed t-test).

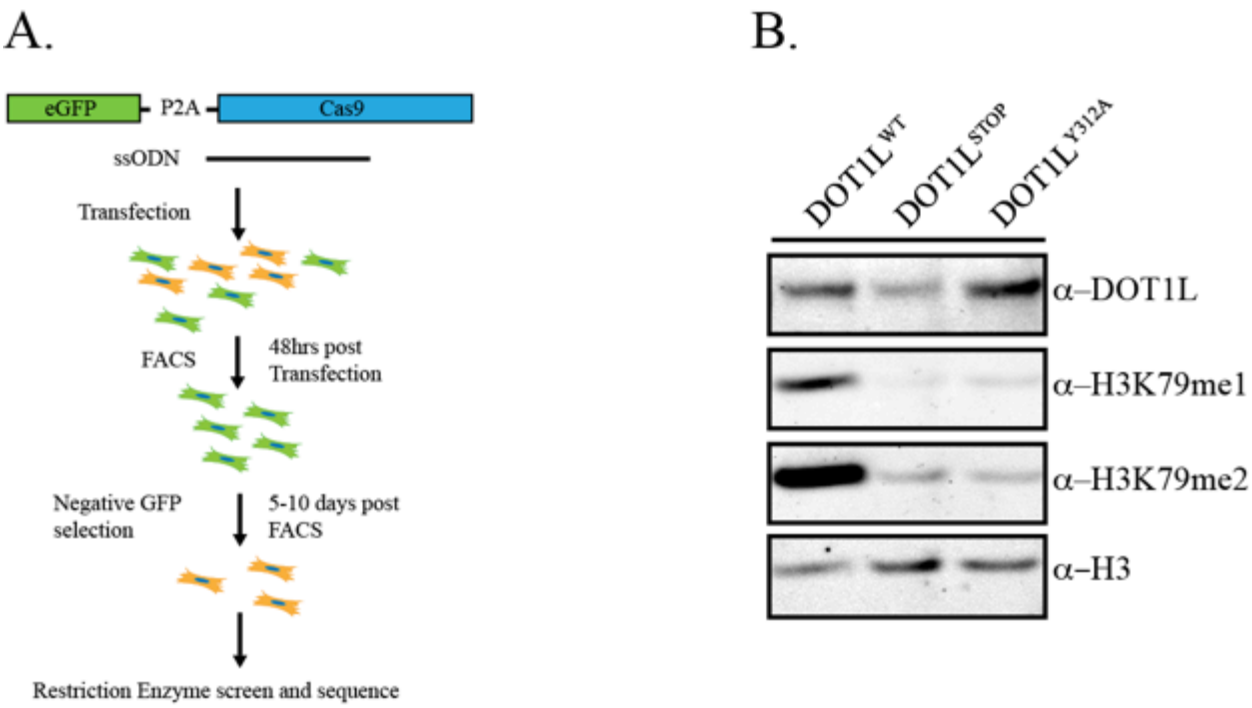


Figure 5.11: Generation of HEK293T DOT1L nonsense and methyltransferase mutants. A. Diagram of Cas9 targeting scheme. eGFP-P2A-Cas9 and a 150nt ssODN are transfected into HEK293T cells and subjected to FACS. Negative eGFP clones were isolated and screened. B. Western of HEK293T clones for DOT1L, H3, H3K79me1, and H3K79me2.

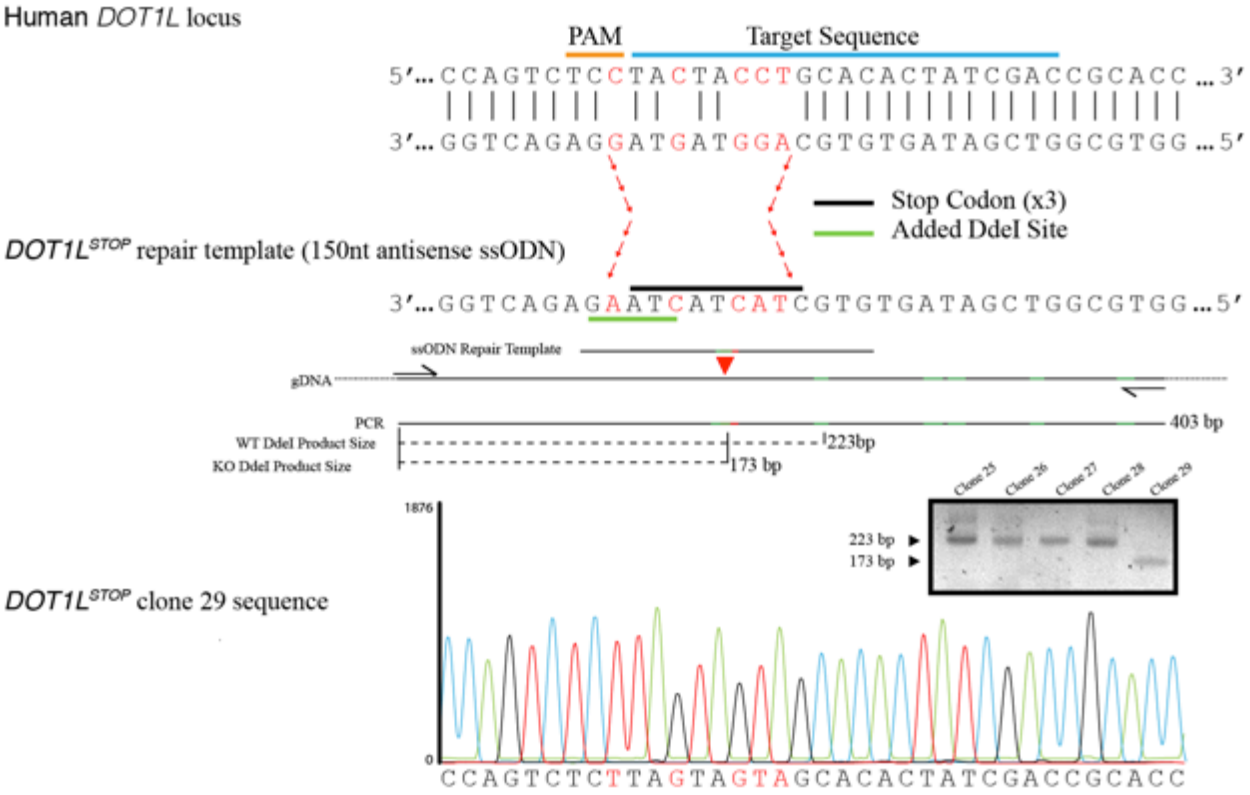


Figure 5.12: Targeting schematic to generate the *DOT1L*^{STOP} HEK293T cell line. Human *DOT1L* locus shows the Cas9 target sequence and PAM site. Red nucleotides show the residues that are targeted for mutation with 150nt ssODN. Green shows the location of the DdeI site that is added. The gel demonstrates the PCR/resection enzyme (DdeI) screening. Bottom is a representative sequence of the region with the mutated residues.

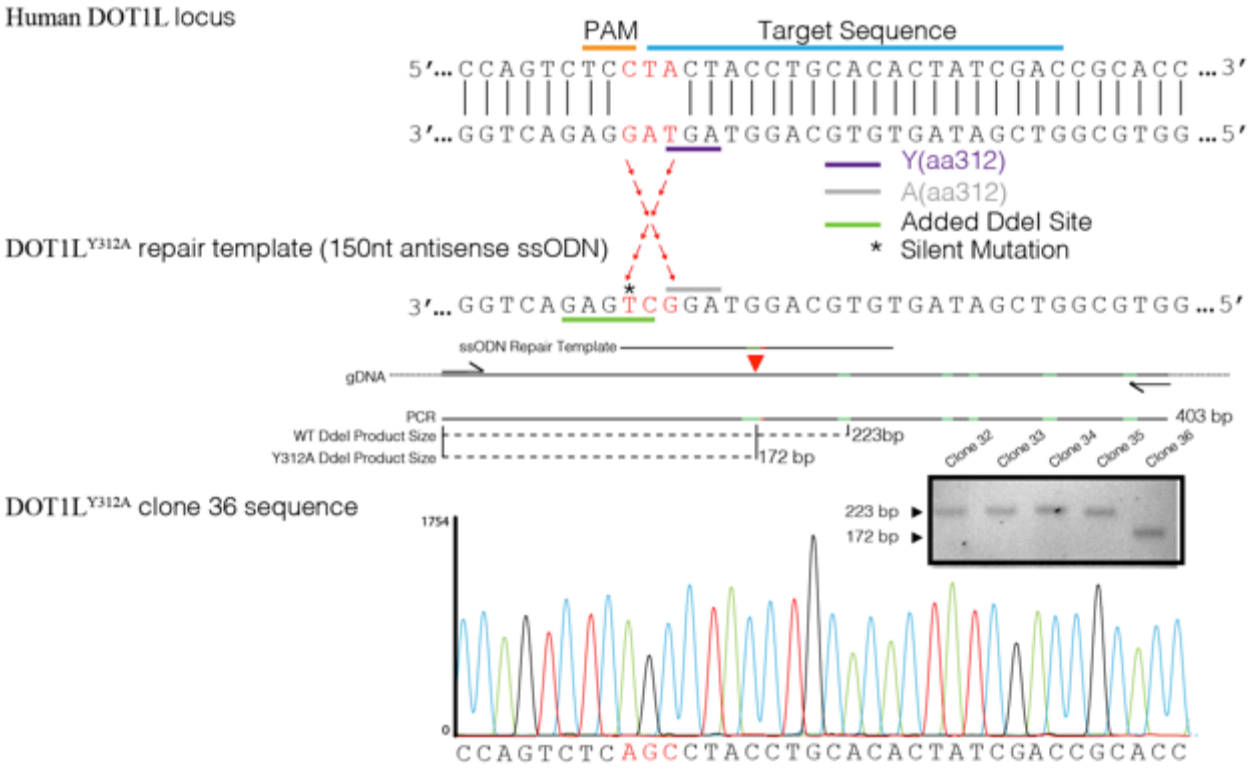


Figure 5.13: Targeting schematic to generate the *DOT1L*^{Y312A} HEK293T cell line. Human *DOT1L* locus shows the Cas9 target sequence and PAM site. Red nucleotides show the residues that are targeted for mutation with 150nt ssODN. Green shows the location of the DdeI site that is added. The gel demonstrates the PCR/resection enzyme (DdeI) screening. Bottom is a representative sequence of the region with the mutated residues.

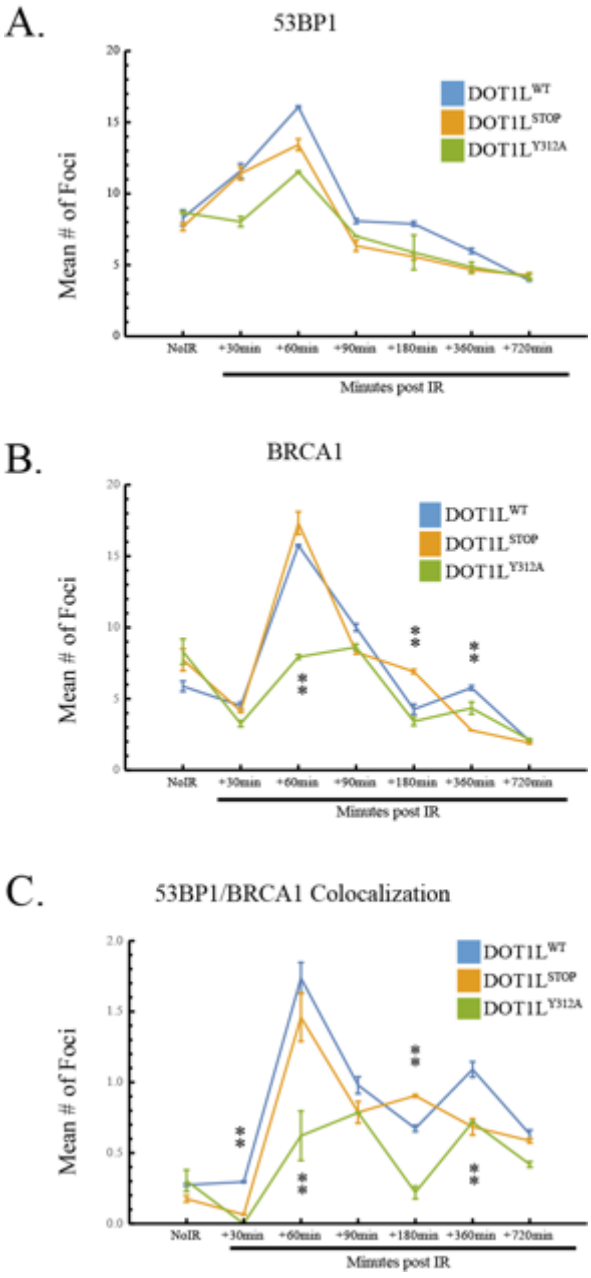


Figure 5.14: *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants have altered kinetics in 53BP1 and BRCA1 foci formation and 53BP1/BRCA1 co localization. DNA damage foci were quantified in non IR treated cells and at 30, 60,90, 180, 360, 720 minutes post IR treatment. A. 53BP1 foci, B. BRCA foci, C. Co-localization of 53BP1/BRCA1 foci whose centroid is within 2 pixel units of each other. ** mark time points with p value < 0.05, pair t-test (wt vs Stop, wt vs Y312a).

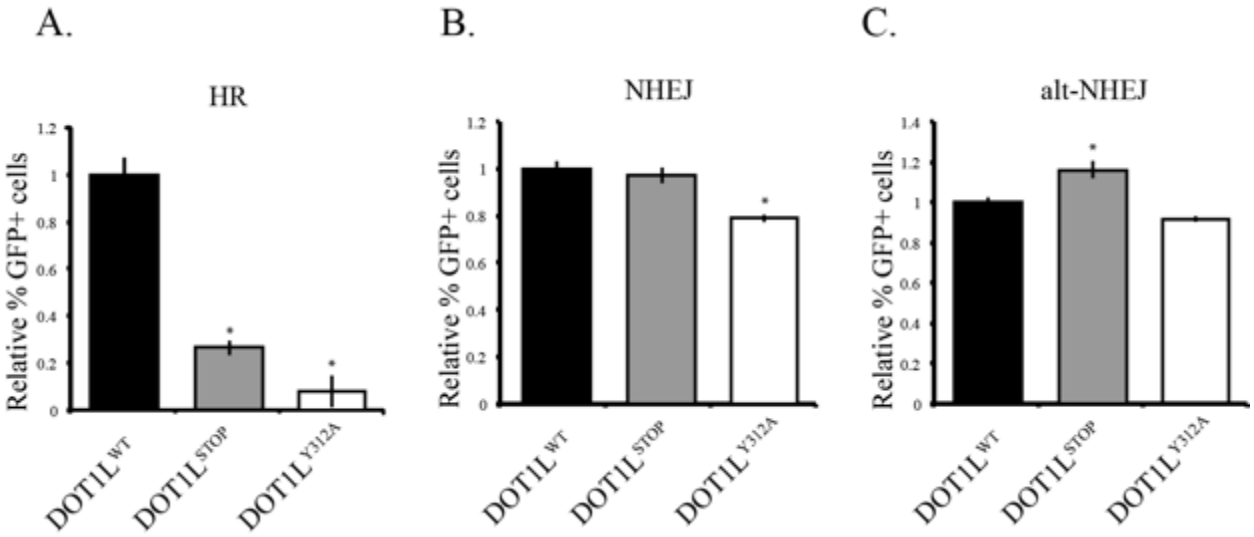


Figure 5.15: *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants have defects in severe defects in HR repair. A. DR-GFP assay was used to monitor HR repair activity in wild-type and *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants. B. E5J-GFP assay was used to monitor NHEJ repair activity in wild-type and *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants. C. E2J-GFP assay was used to monitor HR repair activity in wild-type and *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants. GFP values were normalized to wild-type samples. (A. *DOT1L*^{WT} n=3, *DOT1L*^{STOP}n=3, *DOT1L*^{Y312A} n=3, B. *DOT1L*^{WT} n=3, *DOT1L*^{STOP}n=3, *DOT1L*^{Y312A} n=3,p C. *DOT1L*^{WT} n=3, *DOT1L*^{STOP}n=3, *DOT1L*^{Y312A} n=3value < 0.05, two tailed t-test.)

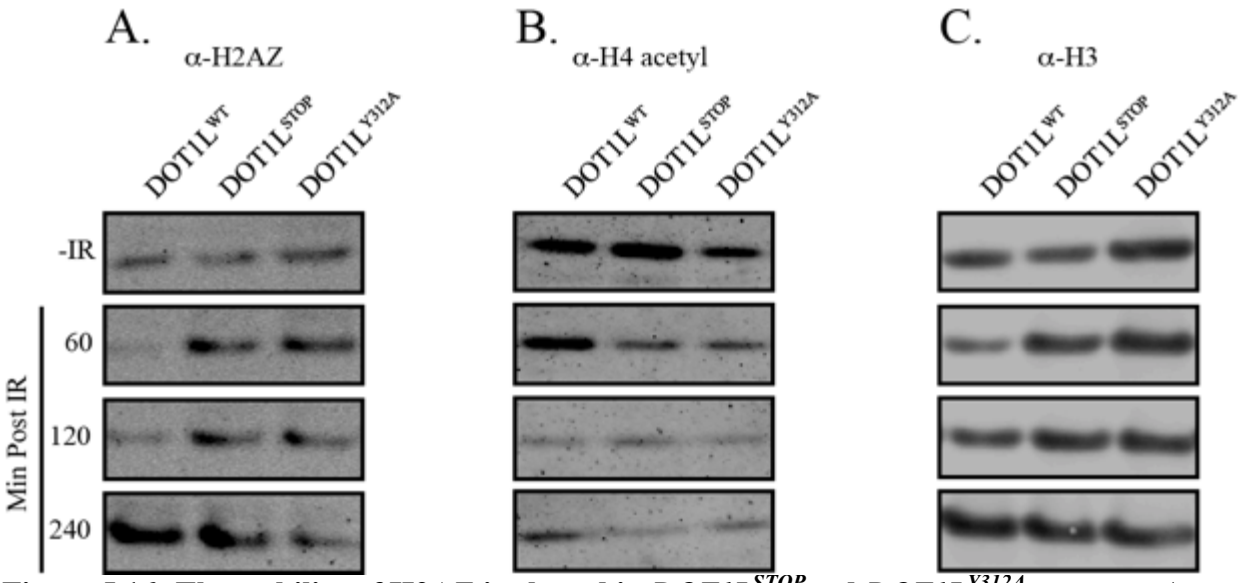


Figure 5.16: The stability of H2AZ is altered in *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants. A. H2AZ histone stability is increased in *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants following treatment with 10 Gy of ionizing radiation. B. H4K acetylation on *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutant histones is reduced. C. H3 histone stability is increased in *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants.

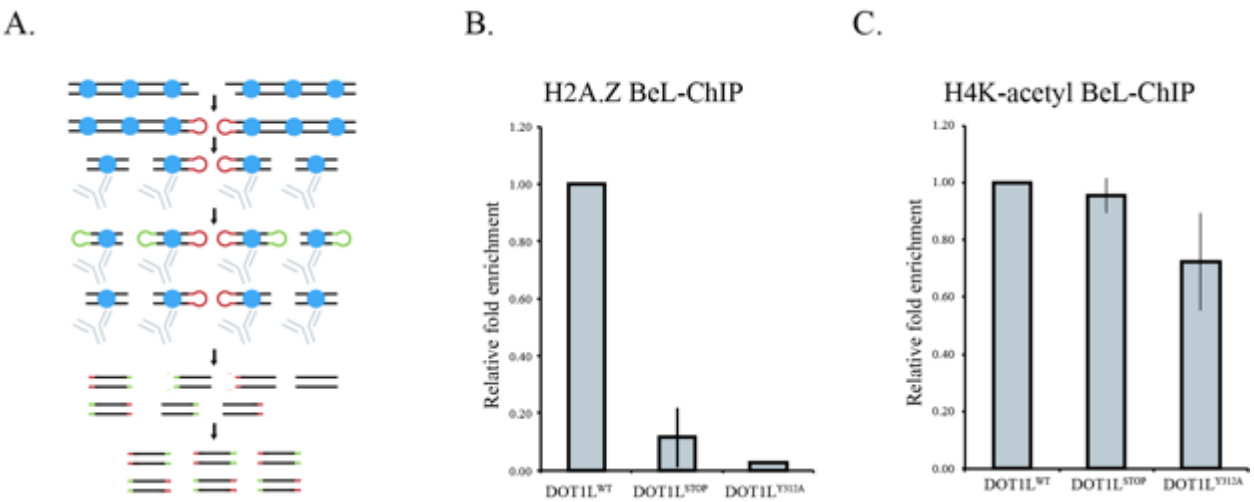


Figure 5.17: H2AZ is lost at the sites of DNA damage. A. Diagram of BeL-Chip assay. Broken end linker (Red) is ligated to the broken end. The genomic DNA is fragmented and crosslinked complexes are immunoprecipitated. A anchor linker (green) is ligated and DNA fragments are purified for analysis. B. qPCR of BeL-ChIP using H2AZ on wildtype, *DOT1L*^{STOP}and *DOT1L*^{Y312A} mutants 60 minutes post ionizing radiation (10Gy) treatment. C. qPCR BeL-ChIP using H4K acetylation on wildtype, *DOT1L*^{STOP}and *DOT1L*^{Y312A} mutants 60 minutes post ionizing radiation (10Gy) treatment. (wildtype n=3, *DOT1L*^{STOP} n=3, *DOT1L*^{Y312A} n=3)

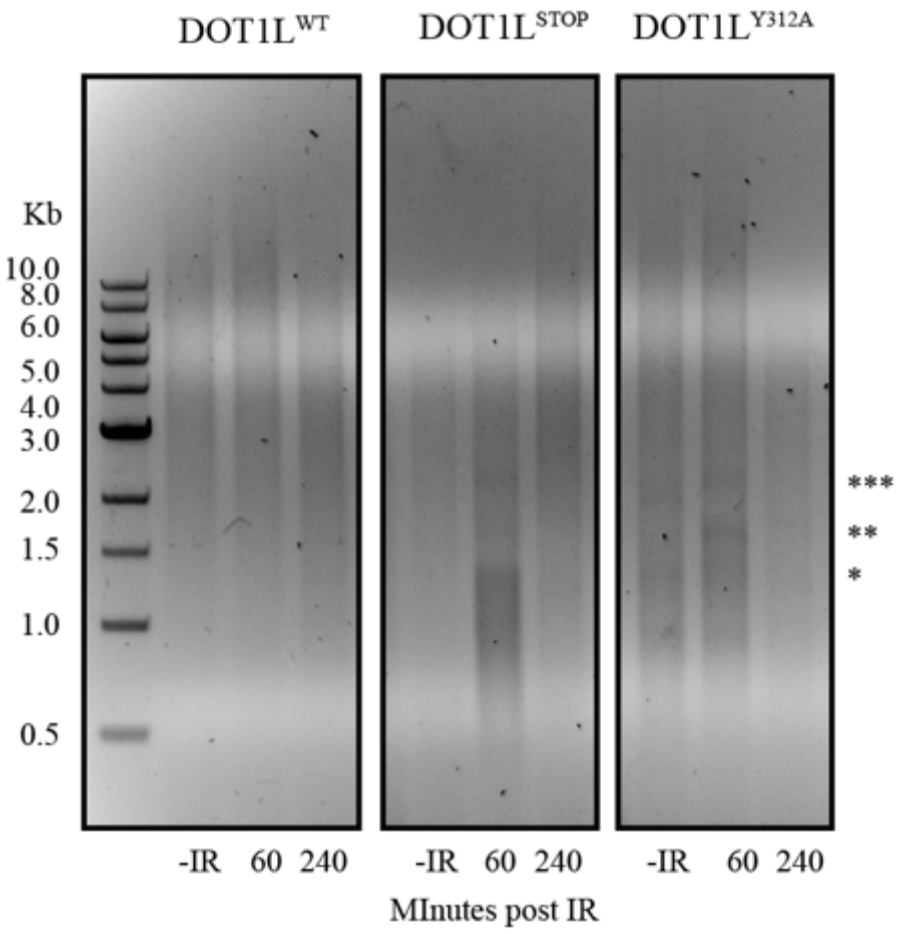


Figure 5.18: *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants have altered single stranded DNA kinetics after DNA damage. Wildtype, *DOT1L*^{STOP} and *DOT1L*^{Y312A} mutants were exposed to 10 Gy of ionizing radiation and ssDNA was isolated at 60 minutes and 240 minutes post ionization radiation. Asterisks mark major species of ssDNA.

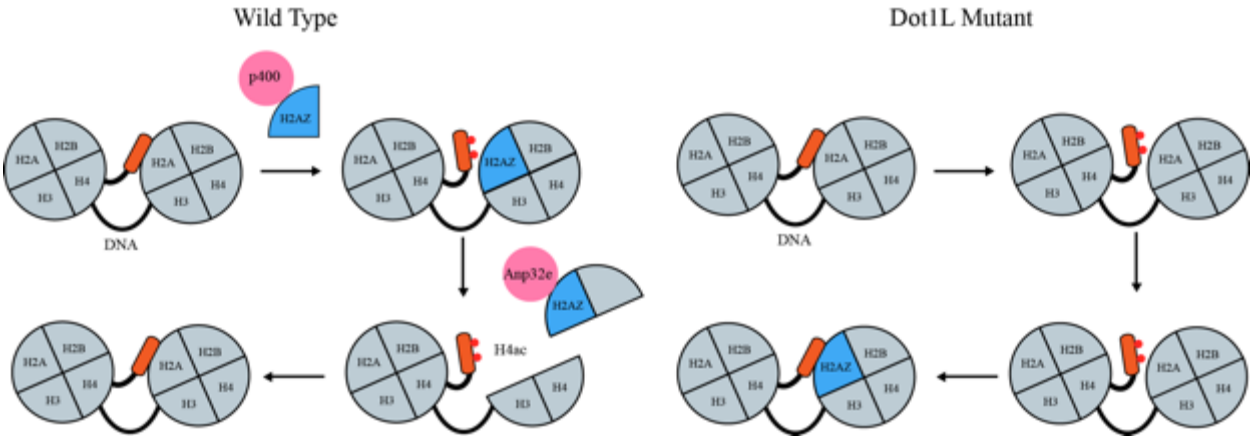


Figure 5.19: Model for *DOT1L* role in H2AZ exchange at sites of DNA damage. Under normal conditions DOT1L facilitates exchange of H2AZ at sites of DNA damage. In the absence of DOT1L H2AZ cannot be exchanged at sites of DNA damage. Our data indicate that H2AZ is more stable in the genome following ionizing radiation treatment in *DOT1L* mutants. This leads us to hypothesize that a lack of H3K79me compacts the chromatin reducing the efficiency of H2AZ exchange. Figure is adopted from the model put forth in [144].

Chapter VI: Concluding remarks and future perspectives.

DNA Damage repair and disease

The fidelity of DNA damage repair pathways is essential for maintaining genome integrity. The loss of DNA repair fidelity will lead to the accumulation of unrepaired DNA lesions that can either lead to cellular death or disease. Spontaneous DNA lesions that escape the DNA repair machinery can be the source of spontaneous cancers and disorders that have no known etiology. However, a significant amount of human maladies is caused by known genetic mutations that alter DNA repair pathway genes. The following section contains examples of DNA repair pathway components and diseases associated with their defects.

Diseases associated with BER

Components of the base excision repair pathway (BER) have been implicated in development of carcinomas and neurological diseases. It was found that OGG1 was directly involved in tumorigenesis in animal models. OGG1 mutations predisposed mice for the development of lung and ovarian tumors, and lymphomas [203]. In humans, missense mutations associated with OGG1 have been linked to kidney carcinomas [204]. The genetic cytology of Huntington disorder, a neurological diseases, is marked by the expansion of CAG nucleotides which is mediated by the age related hyper activity of OGG1[205].

OGG1 activity is partially dependent on the removal of histone 1 (H1) by nucleosome assembly protein 1 (NAP1) [17]. In the case of Huntington disorder, it is conceivable that drugs targeting the epigenetic pathways that facilitate H1 removal could work to slow the progression of the disease by limiting activity of OGG1.

Diseases associated with MMR

Components of the mismatch repair pathway (MMR) have been implicated in development of various carcinoma including colon cancer and leukemias. The loss of functioning mismatch repair can be a result of mutations that silence the MMR genes or point mutations that alter the function of MMR genes. In humans, mutations that occur in the MSH2 gene are associated with Lynch syndrome, a subtype of colorectal cancer [206]. Mutations of the MSH2 and MLH1 have also been linked to cases of T-cell leukemia [207].

Diseases with nucleotide excision repair factors

Nucleotide excision repair (NER) removes diverse DNA lesions, such as after UV-induced damage, bulky chemical adducts, or oxidative damage. Loss of NER proteins results in rare recessive syndromes: xeroderma pigmentosum, Cockayne syndrome, and a photosensitive disorder involving brittle hair disorder trichothiodystrophy [208]. Mutations in the XP genes (XPA,B,C) can cause xeroderma pigmentosum, which is a UV sensitivity that is associated with a more than 1000-fold increased risk of skin carcinoma, stressing the importance of these factors in the repair of UV damage.

Diseases with ICL repair factors

Defects in the repair components of the interstrand crosslinking repair (ICL) pathway predispose humans to a group of heterogeneous disorders that result from genomic instability: developmental abnormalities of major organ systems, bone marrow failure, and a high predisposition to cancer. Cellularly, there is a hypersensitivity to DNA crosslinking agents. Genes commonly associated with these disorders occur in the Fanconi anemia genes: FANCA, B, C, G, F, L [209].

Diseases with SSB repair factors

Single strand break (SSB) repair defects can cause neurological disorders and are also associated with cancers. Mutations in XRCC1 can lead to a neurological disorder known as ataxia-oculomotor apraxia1 (AOA1), which is a type of spinocerebellar ataxia syndrome that includes symptoms such as cognitive impairment, hypercholesterolaemia, and involuntary movements [9]. The neurological manifestation of disease phenotypes associated with SSB repair suggests that the accumulation of SSB, which are the most common lesions in cells, result in progressive neurological dysfunction. Whether defects in SSB repair are associated with neurological degeneration involved in aging remains unknown.

Diseases with DSB repair factors

Double strand break (DSB) repair is central to the maintenance of genomic integrity, therefore it is not surprising that defects in genes central to HR and NHEJ result in cancers and Fanconi anemia. For example, mutations in BRCA1 and BRCA2 predispose women to breast and ovarian cancers [210]. Another example of DSB repair defects occurs in LIG4, the ligase necessary for repair during NHEJ. Hypomorphic mutations of LIG4 result in a syndrome characterized by microcephaly, immunodeficiency, and developmental retardation [211]. Many mutations of DSB repair machinery occur as hypomorphic mutations, suggesting that complete loss of the repair factor would be incompatible with viability.

DOT1L inhibitors for the treatment of leukemia

In the United States there are approximately 13 million individuals are currently living with cancer [212]. There are an estimated 1.6 million new cancer diagnoses each year with approximately 500,000 cancer-related deaths annually. Leukemia, a cancer of the blood, is the

most commonly diagnosed cancer in people aged 20 years and younger, representing approximately 26 percent of all cancer diagnoses in adolescents [212].

A subtype of leukemia termed AML is linked with the lowest survival rates and the leading cause of leukemic deaths. AML patients undergo bone marrow failure due to a loss of hematopoietic progenitors, which decreases the production of neutrophils, platelets, and erythrocytes [213]. AML is characterized by the numerous cytogenetic chromosomal translocations: -5, -7, del(5q), t(6:9), t(9:22), abn17p, abn3q, and abn11q23 [164]. AML with the abn11q23 translocation has been termed Mixed Lineage Leukemia (MLL) and has proven to be one of the most treatment-resistant forms of leukemia. Initial genetic characterization of MLL identified a chromosomal break point at 11q23 causing the lysine methyltransferase 2 (*MLL)* to fuse with the gene MLLT1, super elongation complex subunit (*ENL),* which encodes a transcription elongation factor [166, 167]. Currently, over 100 translocation events have been identified in *MLL* patients, generating over 60 MLL fusion proteins, the most common being AF4/FMR2 family member 1 (AF4), MLLT3, super elongation complex subunit (AF9), ENL, myeloid/lymphoid or mixed-lineage leukemia; translocated to, 10 (AF10), afadin, adherens junction formation factor (AF6), elongation factor for RNA polymerase II (ELL), epidermal growth factor receptor pathway substrate 15 (AF1P), septin 6 (SEPT6), and MLLT6, PHD finger domain containing (AF17) [165]. Of those, AF9, AF10, ENL, AF17 partner with the H3K79 methyltransferase disruptor of telomeric silencing 1-Like (DOT1L), and are part of the DotCom complex, which is involved in transcriptional activation in eukaryotes [108].

A common strategy for drug discovery and anti-hematological cancer therapies is the targeting of epigenetic factors [214]; however, the molecular function of epigenetic regulators must be thoroughly examined in order to improve fficacy. *In vitro* experiments using cell lines

with *MLL* translocations have demonstrated that inhibition of DOT1L methylation can reduce cell viability [215]. We provide evidence here that the cytotoxic effects observed after inhibiting DOT1L activity could be a consequence of its role in DNA repair. Further work is needed in order to gain a detailed understanding of the cellular functions of DOT1L as to improve the efficacy of DOT1L inhibitors.

Final thoughts and future directions

Anti-cancer therapies that exploit DNA repair pathway defects can work to further destabilize the genomes of cancerous cells driving them towards apoptosis. A fundamental understanding of the molecular and cellular functions of DNA repair factors will allow for more specified and improved targeted therapies. In order for these types of molecular therapies to come to fruition, an understanding of the molecular mechanism of DNA repair has to extend beyond cursory observations of end point repair assays. Better methodologies are necessary to gain deeper insights into dynamic processes of DNA repair. Thus, we developed novel and cutting edge methodologies and technologies to study mechanisms of DNA repair. We have applied those methodologies to identify the molecular functions of DOT1L in regulating DNA repair. We envision that the methodologies and technologies developed here can be can be applied to biological processes other than DNA repair factors (i.e. quantification of germ cells in the ovary or neuronal circuit mapping).

In this dissertation we identified the molecular mechanism of the role of DOT1L in DNA double strand break repair. We found that DOT1L regulates histone exchange at sites of DNA double strand breaks. DOT1L is primarily studied as a transcription regulator; however, we found that DOT1L and H3K79 methylation promote the exchange of H2AZ at DSB sites. The exchange of H2AZ into the nucleosome at DSB sites limits the amount of ssDNA generated,

which, if left unchecked, would result in the utilization of highly mutagenic repair pathways. These findings have defined the molecular mechanism for how DOT1L facilitates DNA damage repair and provide insights into how the epigenetic code can direct DNA repair pathway utilization.

In future studies, it will be interesting to examine if other epigenetic factors regulate DNA damage in a similar fashion as DOT1L in double strand break repair, and if the function of DOT1L in DSB repair can be extended to other types of DNA damage repair. DNA damage pathway utilization is influenced by the epigenetic state of the DNA where the damage takes place. Examination of the epigenetic state during different types of DNA repair might reveal precise mechanisms of regulation necessary for damage detection and repair. This information could be used for the development and application of novel molecular therapeutics to treat diseases states or for therapeutic interventions designed to limit the effects of genotoxic environmental factors.

To test our central hypothesis, we developed methodologies and technologies that facilitated in elucidating the molecular mechanism of DOT1L action during DSB DNA repair. The development of BEL-ChIP allows for precise temporal interrogation of nucleosome composition and modifications at sites of DNA damage on a genome wide scale. The nucleosome composition and modification state can influence end resection at DSB sites, which in turn can influence the type of DNA repair pathway used. Next, we developed a method to measure genome wide end resection events at nucleotide resolution using Oxford Nanopore technology. Using this method it is now possible capture all single stranded regions generated after DSB formation. Finally, we developed image analysis software called MANA. MANA can autonomously quantify the spatial and temporal formation of nuclear foci events as well as cell

cycle staging using deep learning neural networks. The usefulness of these technologies and methodologies extend beyond their application in the interrogation of the function of DOT1L in DNA repair and can be applied to interrogation of all epigenetic factors during DNA repair.

Understanding the role of DOT1L during DNA double strand break repair provides greater understanding of how DNA repair is regulated and expands our understanding of how epigenetics can regulate processes beyond transcriptional control. The identification of DOT1L function in DNA repair provides new opportunities for design and implementation of molecular therapeutics strategies that target DNA repair pathways to treat diseases.

Chapter VII: References

1. Swenberg, J.A., et al., *Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment.* Toxicol Sci, 2011. **120 Suppl 1**: p. S130-45.
2. McCulloch, S.D. and T.A. Kunkel, *The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases.* Cell Res, 2008. **18**(1): p. 148-61.
3. Preston, B.D., T.M. Albertson, and A.J. Herr, *DNA replication fidelity and cancer.* Semin Cancer Biol, 2010. **20**(5): p. 281-93.
4. Mullaart, E., et al., *DNA damage metabolism and aging.* Mutation Research/DNAging, 1990. **237**(5-6): p. 189-210.
5. Chaudhry, M.A., *Base excision repair of ionizing radiation-induced DNA damage in G1 and G2 cell cycle phases.* Cancer Cell Int, 2007. **7**: p. 15.
6. Schroering, A.G., et al., *The cell cycle and DNA mismatch repair.* Exp Cell Res, 2007. **313**(2): p. 292-304.
7. Bergink, S., et al., *DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A.* Genes Dev, 2006. **20**(10): p. 1343-52.
8. Mladenova, V. and G. Russev, *Enhanced repair of DNA interstrand crosslinks in S phase.* FEBS Lett, 2006. **580**(6): p. 1631-4.
9. Caldecott, K.W., *Single-strand break repair and genetic disease.* Nat Rev Genet, 2008. **9**(8): p. 619-31.
10. Mao, Z., et al., *DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells.* Cell Cycle, 2008. **7**(18): p. 2902-6.
11. Krokan, H.E. and M. Bjoras, *Base excision repair.* Cold Spring Harb Perspect Biol, 2013. **5**(4): p. a012583.
12. Blainey, P.C., et al., *A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5752-7.
13. Marenstein, D.R., D.M. Wilson, 3rd, and G.W. Teebor, *Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded DNA.* DNA Repair (Amst), 2004. **3**(5): p. 527-33.
14. Fortini, P., et al., *The type of DNA glycosylase determines the base excision repair pathway in mammalian cells.* J Biol Chem, 1999. **274**(21): p. 15230-6.
15. Hanssen-Bauer, A., et al., *XRCC1 coordinates disparate responses and multiprotein repair complexes depending on the nature and context of the DNA damage.* Environ Mol Mutagen, 2011. **52**(8): p. 623-35.
16. Svilar, D., et al., *Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage.* Antioxid Redox Signal, 2011. **14**(12): p. 2491-507.
17. Menoni, H., et al., *Base excision repair of 8-oxoG in dinucleosomes.* Nucleic Acids Res, 2012. **40**(2): p. 692-700.
18. Nakanishi, S., et al., *Different structural states in oligonucleosomes are required for early versus late steps of base excision repair.* Nucleic Acids Res, 2007. **35**(13): p. 4313-21.
19. Kunkel, T.A. and D.A. Erie, *Eukaryotic Mismatch Repair in Relation to DNA Replication.* Annu Rev Genet, 2015. **49**: p. 291-313.

20. Kunkel, T.A., *Nucleotide repeats. Slippery DNA and diseases.* Nature, 1993. **365**(6443): p. 207-8.
21. fJiricny, J., *Mediating mismatch repair.* Nat Genet, 2000. **24**(1): p. 6-8.
22. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.
23. Doerks, T., et al., *Systematic identification of novel protein domain families associated with nuclear functions.* Genome Res, 2002. **12**(1): p. 47-56.
24. Genschel, J., L.R. Bazemore, and P. Modrich, *Human exonuclease I is required for 5' and 3' mismatch repair.* J Biol Chem, 2002. **277**(15): p. 13302-11.
25. Constantin, N., et al., *Human mismatch repair: reconstitution of a nick-directed bidirectional reaction.* J Biol Chem, 2005. **280**(48): p. 39752-61.
26. Jiricny, J., *The multifaceted mismatch-repair system.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.
27. Javaid, S., et al., *Nucleosome remodeling by hMSH2-hMSH6.* Mol Cell, 2009. **36**(6): p. 1086-94.
28. Li, F., et al., *The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha.* Cell, 2013. **153**(3): p. 590-600.
29. Marteijn, J.A., et al., *Understanding nucleotide excision repair and its roles in cancer and ageing.* Nat Rev Mol Cell Biol, 2014. **15**(7): p. 465-81.
30. Scrima, A., et al., *Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex.* Cell, 2008. **135**(7): p. 1213-23.
31. Fousteri, M., et al., *Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo.* Mol Cell, 2006. **23**(4): p. 471-82.
32. Schwertman, P., et al., *UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair.* Nat Genet, 2012. **44**(5): p. 598-602.
33. Zhao, Q., et al., *Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex.* J Biol Chem, 2009. **284**(44): p. 30424-32.
34. Jiang, Y., et al., *INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway.* Proc Natl Acad Sci U S A, 2010. **107**(40): p. 17274-9.
35. Oksenych, V., et al., *Histone methyltransferase DOT1L drives recovery of gene expression after a genotoxic attack.* PLoS Genet, 2013. **9**(7): p. e1003611.
36. Walden, H. and A.J. Deans, *The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder.* Annu Rev Biophys, 2014. **43**: p. 257-78.
37. Ciccia, A., et al., *Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM.* Mol Cell, 2007. **25**(3): p. 331-43.
38. Singh, T.R., et al., *MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM.* Mol Cell, 2010. **37**(6): p. 879-86.
39. Yan, Z., et al., *A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability.* Mol Cell, 2010. **37**(6): p. 865-78.
40. Ulrich, H.D. and H. Walden, *Ubiquitin signalling in DNA replication and repair.* Nat Rev Mol Cell Biol, 2010. **11**(7): p. 479-89.
41. Ishiai, M., et al., *FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway.* Nat Struct Mol Biol, 2008. **15**(11): p. 1138-46.

42. Taniguchi, T., et al., *Convergence of the Fanconi Anemia and Ataxia Telangiectasia Signaling Pathways.* Cell, 2002. **109**(4): p. 459-472.
43. Clauson, C., O.D. Scharer, and L. Niedernhofer, *Advances in understanding the complex mechanisms of DNA interstrand cross-link repair.* Cold Spring Harb Perspect Biol, 2013. **5**(10): p. a012732.
44. Sato, K., et al., *Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair.* EMBO J, 2012. **31**(17): p. 3524-36.
45. Howlett, N., P. Azzinaro, and K. Paquin, *Characterization of a Putative Fanconi Anemia D2 Protein Histone-Binding Domain and Chromodomain.* The FASEB Journal, 2015. **29**(1 Supplement).
46. Eustermann, S., et al., *Structural Basis of Detection and Signaling of DNA Single-Strand Breaks by Human PARP-1.* Mol Cell, 2015. **60**(5): p. 742-54.
47. Muthurajan, U.M., et al., *Automodification switches PARP-1 function from chromatin architectural protein to histone chaperone.* Proc Natl Acad Sci U S A, 2014. **111**(35): p. 12752-7.
48. Gibbs-Seymour, I., et al., *HPF1/C4orf27 Is a PARP-1-Interacting Protein that Regulates PARP-1 ADP-Ribosylation Activity.* Mol Cell, 2016. **62**(3): p. 432-42.
49. Mehrotra, P.V., et al., *DNA repair factor APLF is a histone chaperone.* Mol Cell, 2011. **41**(1): p. 46-55.
50. Iles, N., et al., *APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks.* Mol Cell Biol, 2007. **27**(10): p. 3793-803.
51. Dianova, II, et al., *XRCC1-DNA polymerase beta interaction is required for efficient base excision repair.* Nucleic Acids Res, 2004. **32**(8): p. 2550-5.
52. Wang, J.C., *Cellular roles of DNA topoisomerases: a molecular perspective.* Nat Rev Mol Cell Biol, 2002. **3**(6): p. 430-40.
53. Pourquier, P. and Y. Pommier, *Topoisomerase I-mediated DNA damage.* Adv Cancer Res, 2001. **80**: p. 189-216.
54. Debethune, L., et al., *Processing of nucleopeptides mimicking the topoisomerase I-DNA covalent complex by tyrosyl-DNA phosphodiesterase.* Nucleic Acids Research, 2002. **30**(5): p. 1198-1204.
55. Plo, I., et al., *Association of XRCC1 and tyrosyl DNA phosphodiesterase (Tdp1) for the repair of topoisomerase I-mediated DNA lesions.* DNA Repair, 2003. **2**(10): p. 1087-1100.
56. Messner, S., et al., *PARP1 ADP-ribosylates lysine residues of the core histone tails.* Nucleic Acids Res, 2010. **38**(19): p. 6350-62.
57. Tang, J., et al., *Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination.* Nat Struct Mol Biol, 2013. **20**(3): p. 317-25.
58. Blöcher, D., *DNA Double Strand Breaks in Ehrlich Ascites Tumour Cells at Low Doses of X-rays. I. Determination of Induced Breaks by Centrifugation at Reduced Speed.* International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine, 1982. **42**(3): p. 317-328.
59. Cornforth, M. and J. Bedford, *On the nature of a defect in cells from individuals with ataxia-telangiectasia.* Science, 1985. **227**(4694): p. 1589-1591.

60. Frankenberg-Schwager, M., *Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells.* Radiation and Environmental Biophysics, 1990. **29**(4): p. 273-292.
61. Weibezahn, K.F. and T. Coquerelle, *Radiation induced DNA double strand breaks are rejoined by ligation and recombination processes.* Nucleic Acids Research, 1981. **9**(13): p. 3139-3150.
62. Ali, A.A., et al., *The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks.* Nat Struct Mol Biol, 2012. **19**(7): p. 685-92.
63. Langelier, M.F., et al., *Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1.* Science, 2012. **336**(6082): p. 728-32.
64. Haince, J.F., et al., *PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites.* J Biol Chem, 2008. **283**(2): p. 1197-208.
65. Zhang, F., et al., *The oligonucleotide/oligosaccharide-binding fold motif is a poly(ADP-ribose)-binding domain that mediates DNA damage response.* Proc Natl Acad Sci U S A, 2014. **111**(20): p. 7278-83.
66. Richard, D.J., et al., *hSSB1 rapidly binds at the sites of DNA double-strand breaks and is required for the efficient recruitment of the MRN complex.* Nucleic Acids Res, 2011. **39**(5): p. 1692-702.
67. Stewart, G.S., et al., *MDC1 is a mediator of the mammalian DNA damage checkpoint.* Nature, 2003. **421**(6926): p. 961-966.
68. Munoz, M.C., et al., *RING finger nuclear factor RNF168 is important for defects in homologous recombination caused by loss of the breast cancer susceptibility factor BRCA1.* J Biol Chem, 2012. **287**(48): p. 40618-28.
69. Bennardo, N., et al., *Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair.* PLoS Genet, 2008. **4**(6): p. e1000110.
70. Huen, M.S., et al., *RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly.* Cell, 2007. **131**(5): p. 901-14.
71. Watanabe, S., et al., *JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks.* Nat Struct Mol Biol, 2013. **20**(12): p. 1425-33.
72. Coleman, K.A. and R.A. Greenberg, *The BRCA1-RAP80 complex regulates DNA repair mechanism utilization by restricting end resection.* J Biol Chem, 2011. **286**(15): p. 13669-80.
73. Hu, Y., et al., *RAP80-directed tuning of BRCA1 homologous recombination function at ionizing radiation-induced nuclear foci.* Genes Dev, 2011. **25**(7): p. 685-700.
74. Thorslund, T., et al., *Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage.* Nature, 2015. **527**(7578): p. 389-93.
75. Mallette, F.A., et al., *RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites.* EMBO J, 2012. **31**(8): p. 1865-78.
76. Doil, C., et al., *RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins.* Cell, 2009. **136**(3): p. 435-46.
77. Ochs, F., et al., *53BP1 fosters fidelity of homology-directed DNA repair.* Nat Struct Mol Biol, 2016. **23**(8): p. 714-21.
78. Seroussi, E. and S. Lavi, *Replication protein A is the major single-stranded DNA binding protein detected in mammalian cell extracts by gel retardation assays and UV*

cross-linking of long and short single-stranded DNA molecules. J Biol Chem, 1993. **268**(10): p. 7147-7154.
79. Moore, S.P., et al., *The human homologous pairing protein HPP-1 is specifically stimulated by the cognate single-stranded binding protein hRP-A.* Proceedings of the National Academy of Sciences, 1991. **88**(20): p. 9067-9071.
80. Iftode, C. and J.A. Borowiec, *5′→3′ Molecular Polarity of Human Replication Protein A (hRPA) Binding to Pseudo-Origin DNA Substrates †.* Biochemistry, 2000. **39**(39): p. 11970-11981.
81. Chen, J., et al., *Mechanochemical regulations of RPA's binding to ssDNA.* Sci Rep, 2015. **5**: p. 9296.
82. Grimme, J.M., et al., *Human Rad52 binds and wraps single-stranded DNA and mediates annealing via two hRad52-ssDNA complexes.* Nucleic Acids Res, 2010. **38**(9): p. 2917-30.
83. Qi, W., et al., *BRG1 promotes the repair of DNA double-strand breaks by facilitating the replacement of RPA with RAD51.* J Cell Sci, 2015. **128**(2): p. 317-30.
84. Zhao, W., et al., *Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry.* Mol Cell, 2015. **59**(2): p. 176-87.
85. McIlwraith, M.J., et al., *Reconstitution of the strand invasion step of double-strand break repair using human Rad51 Rad52 and RPA proteins.* J Mol Biol, 2000. **304**(2): p. 151-64.
86. Svendsen, J.M., et al., *Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair.* Cell, 2009. **138**(1): p. 63-77.
87. Yun, M.H. and K. Hiom, *CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle.* Nature, 2009. **459**(7245): p. 460-3.
88. Sartori, A.A., et al., *Human CtIP promotes DNA end resection.* Nature, 2007. **450**(7169): p. 509-14.
89. Chen, L., et al., *Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair.* J Biol Chem, 2008. **283**(12): p. 7713-20.
90. Simsek, D. and M. Jasin, *Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation.* Nat Struct Mol Biol, 2010. **17**(4): p. 410-6.
91. Deng, S.K., et al., *RPA antagonizes microhomology-mediated repair of DNA double-strand breaks.* Nat Struct Mol Biol, 2014. **21**(4): p. 405-12.
92. Audebert, M., B. Salles, and P. Calsou, *Effect of double-strand break DNA sequence on the PARP-1 NHEJ pathway.* Biochem Biophys Res Commun, 2008. **369**(3): p. 982-8.
93. Audebert, M., B. Salles, and P. Calsou, *Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining.* J Biol Chem, 2004. **279**(53): p. 55117-26.
94. Cook, R., et al., *Direct involvement of retinoblastoma family proteins in DNA repair by non-homologous end-joining.* Cell Rep, 2015. **10**(12): p. 2006-18.
95. Carr, S.M., et al., *Lysine methylation-dependent binding of 53BP1 to the pRb tumor suppressor.* Proc Natl Acad Sci U S A, 2014. **111**(31): p. 11341-6.
96. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annu Rev Biochem, 2010. **79**: p. 181-211.

97. Walker, J.R., R.A. Corpina, and J. Goldberg, *Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair.* Nature, 2001. **412**(6847): p. 607-14.
98. Arosio, D., et al., *Fluorescence anisotropy studies on the Ku-DNA interaction: anion and cation effects.* J Biol Chem, 2004. **279**(41): p. 42826-35.
99. Gagne, J.P., et al., *Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes.* Nucleic Acids Res, 2008. **36**(22): p. 6959-76.
100. Couto, C.A., et al., *PARP regulates nonhomologous end joining through retention of Ku at double-strand breaks.* J Cell Biol, 2011. **194**(3): p. 367-75.
101. Ng, H.H., et al., *Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association.* Genes Dev, 2002. **16**(12): p. 1518-27.
102. Feng, Q., et al., *Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain.* Current Biology, 2002. **12**(12): p. 1052-1058.
103. van Leeuwen, F., P.R. Gafken, and D.E. Gottschling, *Dot1p modulates silencing in yeast by methylation of the nucleosome core.* Cell, 2002. **109**(6): p. 745-56.
104. M S Singer, A.K., A J Wolf, L L Meisinger, S E Peterson, C Goggin, M Mahowald, and D E Gottschling, *Identification of High-Copy Disruptors of Telomeric Silencing in Saccharomyces cerevisiae.* Genetics, 1998. **150**(2).
105. Schubeler, D., et al., *The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote.* Genes Dev, 2004. **18**(11): p. 1263-71.
106. Steger, D.J., et al., *DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells.* Mol Cell Biol, 2008. **28**(8): p. 2825-39.
107. Okada, Y., et al., *hDOT1L links histone methylation to leukemogenesis.* Cell, 2005. **121**(2): p. 167-78.
108. Mohan, M., et al., *Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom).* Genes Dev, 2010. **24**(6): p. 574-89.
109. Chang, M.J., et al., *Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes.* Cancer Res, 2010. **70**(24): p. 10234-42.
110. Daigle, S.R., et al., *Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor.* Cancer Cell, 2011. **20**(1): p. 53-65.
111. aBernt, K.M., et al., *MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L.* Cancer Cell, 2011. **20**(1): p. 66-78.
112. Lazzaro, F., et al., *Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres.* EMBO J, 2008. **27**(10): p. 1502-12.
113. Wysocki, R., et al., *Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9.* Mol Cell Biol, 2005. **25**(19): p. 8430-43.
114. Giannattasio, M., et al., *The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1.* J Biol Chem, 2005. **280**(11): p. 9879-86.
115. Botuyan, M.V., et al., *Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair.* Cell, 2006. **127**(7): p. 1361-73.

116. Wakeman, T.P., et al., *Bat3 facilitates H3K79 dimethylation by DOT1L and promotes DNA damage-induced 53BP1 foci at G1/G2 cell-cycle phases.* EMBO J, 2012. **31**(9): p. 2169-81.
117. San-Segundo, P.A. and G.S. Roeder, *Pch2 links chromatin silencing to meiotic checkpoint control.* Cell, 1999. **97**(3): p. 313-24.
118. Roeder, G.S. and J.M. Bailis, *The pachytene checkpoint.* Trends Genet, 2000. **16**(9): p. 395-403.
119. San-Segundo, P.A. and G.S. Roeder, *Role for the silencing protein Dot1 in meiotic checkpoint control.* Mol Biol Cell, 2000. **11**(10): p. 3601-15.
120. Game, J.C., M.S. Williamson, and C. Baccari, *X-ray survival characteristics and genetic analysis for nine Saccharomyces deletion mutants that show altered radiation sensitivity.* Genetics, 2005. **169**(1): p. 51-63.
121. Briggs, S.D., et al., *Gene silencing: trans-histone regulatory pathway in chromatin.* Nature, 2002. **418**(6897): p. 498.
122. Ng, H.H., et al., *Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79.* J Biol Chem, 2002. **277**(38): p. 34655-7.
123. Wood, A., et al., *Bre1, an E3 Ubiquitin Ligase Required for Recruitment and Substrate Selection of Rad6 at a Promoter.* Molecular Cell, 2003. **11**(1): p. 267-274.
124. Conde, F., et al., *The Dot1 histone methyltransferase and the Rad9 checkpoint adaptor contribute to cohesin-dependent double-strand break repair by sister chromatid recombination in Saccharomyces cerevisiae.* Genetics, 2009. **182**(2): p. 437-46.
125. Chen, X., et al., *The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends.* Nature, 2012. **489**(7417): p. 576-80.
126. Huyen, Y., et al., *Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks.* Nature, 2004. **432**(7015): p. 406-11.
127. FitzGerald, J., et al., *Regulation of the DNA damage response and gene expression by the Dot1L histone methyltransferase and the 53Bp1 tumour suppressor.* PLoS One, 2011. **6**(2): p. e14714.
128. Fradet-Turcotte, A., et al., *53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark.* Nature, 2013. **499**(7456): p. 50-4.
129. Lu, X., et al., *The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure.* Nat Struct Mol Biol, 2008. **15**(10): p. 1122-4.
130. Cho, M.H., et al., *DOT1L cooperates with the c-Myc-p300 complex to epigenetically derepress CDH1 transcription factors in breast cancer progression.* Nat Commun, 2015. **6**: p. 7821.
131. Das, C., et al., *CBP/p300-mediated acetylation of histone H3 on lysine 56.* Nature, 2009. **459**(7243): p. 113-7.
132. Pasini, D., et al., *Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes.* Nucleic Acids Res, 2010. **38**(15): p. 4958-69.
133. Henry, R.A., Y.M. Kuo, and A.J. Andrews, *Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4.* Biochemistry, 2013. **52**(34): p. 5746-59.

134. Dovey, O.M., C.T. Foster, and S.M. Cowley, *Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation.* Proc Natl Acad Sci U S A, 2010. **107**(18): p. 8242-7.
135. Frank, S.R., et al., *MYC recruits the TIP60 histone acetyltransferase complex to chromatin.* EMBO Rep, 2003. **4**(6): p. 575-80.
136. Xu, Y., et al., *Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair.* Mol Cell, 2012. **48**(5): p. 723-33.
137. Xu, Y., et al., *The p400 ATPase regulates nucleosome stability and chromatin ubiquitination during DNA repair.* J Cell Biol, 2010. **191**(1): p. 31-43.
138. Li, Z., et al., *c-Myc Suppression of DNA Double-strand Break Repair.* Neoplasia, 2012. **14**(12): p. 1190-IN35.
139. Vempati, R.K., et al., *p300-mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals.* J Biol Chem, 2010. **285**(37): p. 28553-64.
140. Stulemeijer, I.J., et al., *Dot1 binding induces chromatin rearrangements by histone methylation-dependent and -independent mechanisms.* Epigenetics Chromatin, 2011. **4**(1): p. 2.
141. Crosetto, N., et al., *Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing.* Nat Methods, 2013. **10**(4): p. 361-5.
142. Hornsby, P.J. and V.V. Didenko, *In situ ligation: a decade and a half of experience.* Methods Mol Biol, 2011. **682**: p. 49-63.
143. Didenko VV, B.D., Baskin DS., *Substantial Background Reduction in Ligase-based Apoptosis Detection Using Newly Designed Hairpin Oligoprobes.* Biotechniques, 1999. **27**(6): p. 1130-2.
144. Gursoy-Yuzugullu, O., M.K. Ayrapetov, and B.D. Price, *Histone chaperone Anp32e removes H2A.Z from DNA double-strand breaks and promotes nucleosome reorganization and DNA repair.* Proc Natl Acad Sci U S A, 2015. **112**(24): p. 7507-12.
145. Zhou, Y., et al., *Quantitation of DNA double-strand break resection intermediates in human cells.* Nucleic Acids Res, 2014. **42**(3): p. e19.
146. Cruz-Garcia, A., A. Lopez-Saavedra, and P. Huertas, *BRCA1 accelerates CtIP-mediated DNA-end resection.* Cell Rep, 2014. **9**(2): p. 451-9.
147. Urban, J.M., et al., *Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION.* Preprint at http://biorxiv.
org/content/early/2015/06/22/019281 (2015), 2015.
148. Escribano-Diaz, C., et al., *A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice.* Mol Cell, 2013. **49**(5): p. 872-83.
149. fPapamichos-Chronakis, M. and C.L. Peterson, *Chromatin and the genome integrity network.* Nat Rev Genet, 2013. **14**(1): p. 62-75.
150. Rassool, F.V., et al., *Constitutive DNA damage is linked to DNA replication abnormalities in Bloom's syndrome cells.* Oncogene, 2003. **22**(54): p. 8749-57.
151. Schonn, I., J. Hennesen, and D.C. Dartsch, *Cellular responses to etoposide: cell death despite cell cycle arrest and repair of DNA damage.* Apoptosis, 2010. **15**(2): p. 162-72.
152. Sakaue-Sawano, A., et al., *Visualizing spatiotemporal dynamics of multicellular cell-cycle progression.* Cell, 2008. **132**(3): p. 487-98.
153. Arias, E.E. and J.C. Walter, *PCNA functions as a molecular platform to trigger Cdt1 destruction and prevent re-replication.* Nat Cell Biol, 2006. **8**(1): p. 84-90.

154. Roukos, V., et al., *Dynamic recruitment of licensing factor Cdt1 to sites of DNA damage.* J Cell Sci, 2011. **124**(Pt 3): p. 422-34.
155. Roukos, V., et al., *Cell cycle staging of individual cells by fluorescence microscopy.* Nat Protoc, 2015. **10**(2): p. 334-48.
156. Lucy, L.B., *An iterative technique for the rectification of observed distributions.* The Astronomical Journal, 1974. **79**: p. 745.
157. Richardson, W.H., *Bayesian-Based Iterative Method of Image Restoration*.* Journal of the Optical Society of America, 1972. **62**(1): p. 55.
158. Blasi, T., et al., *Label-free cell cycle analysis for high-throughput imaging flow cytometry.* Nat Commun, 2016. **7**: p. 10256.
159. Watson, J.V., S.H. Chambers, and P.J. Smith, *A pragmatic approach to the analysis of DNA histograms with a definable G1 peak.* Cytometry, 1987. **8**(1): p. 1-8.
160. Chu, Y.L., et al., *MutS homologue hMSH4: interaction with eIF3f and a role in NHEJ-mediated DSB repair.* Mol Cancer, 2013. **12**: p. 51.
161. Downs, J.A., et al., *Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites.* Mol Cell, 2004. **16**(6): p. 979-90.
162. Feng, Y., et al., *Early mammalian erythropoiesis requires the Dot1L methyltransferase.* Blood, 2010. **116**(22): p. 4483-91.
163. Pierce, A.J., et al., *XRCC3 promotes homology-directed repair of DNA damage in mammalian cells.* Genes Dev, 1999. **13**(20): p. 2633-8.
164. Lowenberg, B., *Acute myeloid leukemia: the challenge of capturing disease variety.* Hematology Am Soc Hematol Educ Program, 2008: p. 1-11.
165. Meyer, C., et al., *New insights to the MLL recombinome of acute leukemias.* Leukemia, 2009. **23**(8): p. 1490-9.
166. Mueller, D., et al., *A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification.* Blood, 2007. **110**(13): p. 4445-54.
167. Tkachuk, D.C., S. Kohler, and M.L. Cleary, *Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias.* Cell, 1992. **71**(4): p. 691-700.
168. Beerman, I., et al., *Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle.* Cell Stem Cell, 2014. **15**(1): p. 37-50.
169. Walter, D., et al., *Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells.* Nature, 2015. **520**(7548): p. 549-52.
170. Mohrin, M., et al., *Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis.* Cell Stem Cell, 2010. **7**(2): p. 174-85.
171. Rossi, D.J., et al., *Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age.* Nature, 2007. **447**(7145): p. 725-9.
172. Nijnik, A., et al., *DNA repair is limiting for haematopoietic stem cells during ageing.* Nature, 2007. **447**(7145): p. 686-90.
173. Bender, C.F., et al., *Cancer predisposition and hematopoietic failure in Rad50(S/S) mice.* Genes Dev, 2002. **16**(17): p. 2237-51.
174. Parmar, K., et al., *Hematopoietic stem cell defects in mice with deficiency of Fancd2 or Usp1.* Stem Cells, 2010. **28**(7): p. 1186-95.
175. Jones, B., et al., *The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure.* PLoS Genet, 2008. **4**(9): p. e1000190.

176. Hirschi, K.K., *Hemogenic endothelium during development and beyond.* Blood, 2012. **119**(21): p. 4823-7.
177. Vakifahmetoglu, H., M. Olsson, and B. Zhivotovsky, *Death through a tragedy: mitotic catastrophe.* Cell Death Differ, 2008. **15**(7): p. 1153-62.
178. Crasta, K., et al., *DNA breaks and chromosome pulverization from errors in mitosis.* Nature, 2012. **482**(7383): p. 53-8.
179. Zhang, C.Z., et al., *Chromothripsis from DNA damage in micronuclei.* Nature, 2015. **522**(7555): p. 179-84.
180. Bartek, J., C. Lukas, and J. Lukas, *Checking on DNA damage in S phase.* Nat Rev Mol Cell Biol, 2004. **5**(10): p. 792-804.
181. Ahel, D., et al., *Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1.* Science, 2009. **325**(5945): p. 1240-3.
182. Gottschalk, A.J., et al., *Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13770-4.
183. Satoh, M.S. and T. Lindahl, *Role of poly(ADP-ribose) formation in DNA repair.* Nature, 1992. **356**(6367): p. 356-8.
184. Illuzzi, G., et al., *PARG is dispensable for recovery from transient replicative stress but required to prevent detrimental accumulation of poly(ADP-ribose) upon prolonged replicative stress.* Nucleic Acids Res, 2014. **42**(12): p. 7776-92.
185. Holloman, W.K., *Unraveling the mechanism of BRCA2 in homologous recombination.* Nat Struct Mol Biol, 2011. **18**(7): p. 748-54.
186. Zhu, Z., et al., *Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends.* Cell, 2008. **134**(6): p. 981-94.
187. Cejka, P., et al., *DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2.* Nature, 2010. **467**(7311): p. 112-6.
188. Daley, J.M., et al., *Multifaceted role of the Topo IIIalpha-RMI1-RMI2 complex and DNA2 in the BLM-dependent pathway of DNA break end resection.* Nucleic Acids Res, 2014. **42**(17): p. 11083-91.
189. Ip, S.C., et al., *Identification of Holliday junction resolvases from humans and yeast.* Nature, 2008. **456**(7220): p. 357-61.
190. Wyatt, H.D., et al., *Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells.* Mol Cell, 2013. **52**(2): p. 234-47.
191. Munoz, I.M., et al., *Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair.* Mol Cell, 2009. **35**(1): p. 116-27.
192. Pankotai, T., et al., *DNAPKcs-dependent arrest of RNA polymerase II transcription in the presence of DNA breaks.* Nat Struct Mol Biol, 2012. **19**(3): p. 276-82.
193. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.
194. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
195. Min, J., et al., *Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase.* Cell, 2003. **112**(5): p. 711-23.
196. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

197. Chapman, J.R., et al., *RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection.* Mol Cell, 2013. **49**(5): p. 858-71.
198. Luo, K., et al., *Sumoylation of MDC1 is important for proper DNA damage response.* EMBO J, 2012. **31**(13): p. 3008-19.
199. Lee, J.H., Paull, T.T., *ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex.* Science, 2005. **308**(5721): p. 551-554.
200. Alvarez-Quilon, A., et al., *ATM specifically mediates repair of double-strand breaks with blocked DNA ends.* Nat Commun, 2014. **5**: p. 3347.
201. Panier, S. and S.J. Boulton, *Double-strand break repair: 53BP1 comes into focus.* Nat Rev Mol Cell Biol, 2014. **15**(1): p. 7-18.
202. Zong, D., A.R. Chaudhuri, and A. Nussenzweig, *More end resection is not merrier.* Nat Struct Mol Biol, 2016. **23**(8): p. 699-701.
203. Xie, Y., et al., *Deficiencies in Mouse Myh and Ogg1 Result in Tumor Predisposition and G to T Mutations in Codon 12 of the K-Ras Oncogene in Lung Tumors.* Cancer Research, 2004. **64**(9): p. 3096-3102.
204. Audebert, M., et al., *Alterations of the DNA repair gene OGG1 in human clear cell carcinomas of the kidney.* Cancer Res, 2000. **60**(17): p. 4740-4.
205. Kovtun, I.V., et al., *OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells.* Nature, 2007. **447**(7143): p. 447-52.
206. Lynch, H.T., et al., *Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications.* Clin Genet, 2009. **76**(1): p. 1-18.
207. Whiteside, D., et al., *A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots.* Cancer Res, 2002. **62**(2): p. 359-62.
208. de Boer, J., *Nucleotide excision repair and human syndromes.* Carcinogenesis, 2000. **21**(3): p. 453-460.
209. Deakyne, J.S. and A.V. Mazin, *Fanconi anemia: at the Crossroads of DNA repair.* Biochemistry (Moscow), 2011. **76**(1): p. 36-48.
210. Offit, K., et al., *Shared Genetic Susceptibility to Breast Cancer, Brain Tumors, and Fanconi Anemia.* JNCI Journal of the National Cancer Institute, 2003. **95**(20): p. 1548-1551.
211. O'Driscoll, M., et al., *An overview of three new disorders associated with genetic instability: LIG4 syndrome, RS-SCID and ATR-Seckel syndrome.* DNA Repair (Amst), 2004. **3**(8-9): p. 1227-35.
212. Hayat, M.J., et al., *Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program.* Oncologist, 2007. **12**(1): p. 20-37.
213. Miraki-Moud, F., et al., *Acute myeloid leukemia does not deplete normal hematopoietic stem cells but induces cytopenias by impeding their differentiation.* Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13576-81.
214. Mack, G.S., *To selectivity and beyond.* Nat Biotechnol, 2010. **28**(12): p. 1259-66.
215. Yu, W., et al., *Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors.* Nat Commun, 2012. **3**: p. 1288.